Panel error correction testing with global stochastic trends

Christian Gengenbach, Jean-Pierre Urbain,
Joakim Westerlund

Panel Error Correction Testing with Global Stochastic Trends

RM/08/051

JEL code: C12, C33

Maastricht research school of **E**conomics
of **TE**chnology and **OR**ganizations

Universiteit Maastricht
Faculty of Economics and Business Administration
P.O. Box 616
NL - 6200 MD Maastricht

phone : ++31 43 388 3830
fax : ++31 43 388 4873

Panel Error Correction Testing with Global Stochastic Trends*

Christian Gengenbach
Universiteit Maastricht
The Netherlands

Jean-Pierre Urbain
Universiteit Maastricht
The Netherlands

Joakim Westerlund†
Lund University
Sweden

December 15, 2008

Abstract

This paper considers a cointegrated panel data model with common factors. Starting from the triangular representation of the model as used by Bai *et al.* (2008) a Granger type representation theorem is derived. The conditional error correction representation is obtained, which is used as a basis for developing two new tests for the null hypothesis of no error correction. The asymptotic distributions of the tests are shown to be free of nuisance parameters, depending only on the number of non-stationary variables. However, the tests are not cross-sectionally independent, which makes pooling difficult. Nevertheless, the averages of the tests converge in distribution. This makes pooling possible in spite of the cross-sectional dependence. We investigate the finite sample performance of the proposed tests in a Monte Carlo experiment and compare them to the tests proposed by Westerlund (2007). We also present two empirical applications of the new tests.

Keywords: Panel cointegration, common factors.

*Previous versions of this paper were presented at the 2008 conference on Factor Structures for Panel and Multivariate Time Series Data in Maastricht and at a seminar at Lund University. The author would like to thank conference and seminar participants, and in particular David Edgerton, Maria Holmlund, Barry Jones and Franz Palm for many valuable comments and suggestions. Westerlund would also like to thank the Maastricht Research School of Economics of Technology and Organizations for its hospitality during a visit at the Department of Quantitative Economics at Universiteit Maastricht, where this work was initiated. Financial support from the Jan Wallander and Tom Hedelius Foundation under research grant number W2006–0068:1 is gratefully acknowledged.

†Corresponding author: Department of Economics, Lund University, P. O. Box 7082, SE–220 07 Lund, Sweden. Telephone: +46 46 222 8670, Fax: +46 46 222 4118, E-mail address: `joakim.westerlund@nek.lu.se`.

1 Introduction

Consider two non-stationary panel data variables $X_{i,t}$ and $Y_{i,t}$, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$ indexes the cross-sectional and time series dimensions, respectively. The analysis of such variables has been a growing field of econometric research in recent years. See for example Breitung and Pesaran (2008) for an overview. In particular, in many economic applications it is an important question whether $X_{i,t}$ and $Y_{i,t}$ are cointegrated, that is whether there exists a meaningful long-run relationship between them, or whether the relationship is spurious.

Kao (1999) and Pedroni (1999, 2004) were among the first to propose residual-based tests for the null hypothesis of no cointegration in cross-sectionally independent panels. But cross-sectional independence is a restrictive assumption that is unlikely to be met in practice, in which case the properties of this kind of tests become suspect. In fact, in a recent paper, Gengenbach *et al.* (2006) show that the presence of cross-section dependence in the form of non-stationary common factors can actually cause the residual-based tests of Kao (1999) and Pedroni (2004) to become divergent. As a response to this, they propose to estimate separately the common and idiosyncratic components of $X_{i,t}$ and $Y_{i,t}$ using the principal components method of Bai and Ng (2004), and then to test for cointegration in the resulting component estimates.

Banerjee and Carrion-i-Silvestre (2006) propose a similar test but instead of applying the Bai and Ng (2004) approach to $X_{i,t}$ and $Y_{i,t}$ directly, they apply it to the residuals of a first-stage regression of $Y_{i,t}$ onto $X_{i,t}$. Cointegration requires that both the common and idiosyncratic components of the residuals are stationary. The tests of Bai and Carrion-i-Silvestre (2007), Westerlund (2007) and Westerlund and Edgerton (2008) are basically the same in the sense that they are also based on applying the Bai and Ng (2004) approach to the residuals of a first-stage regression.

However, although very popular, this testing approach has at least two major drawbacks. One lies with the use of residual rather than structural dynamics, which makes it subject to the common factor critique of Kremers *et al.* (1992), that may lead to tests with low power. The second drawback is that the testing must be carried out in steps, with the estimation error from one step being imported into subsequent steps, and it is not fully clear what effect this has on the final test, see Westerlund and Larsson (2008).

By contrast to the test proposed by Pedroni for example, the tests of Westerlund (2007) are not based on residuals but rather on the significance of the error correction term in a conditional panel error correction model (ECM), and therefore do not impose any common factor restriction. However, the tests are derived under cross-sectional independence, and the use of the bootstrap in case of violations does not fit well with the otherwise parametric flavor of the tests. Another drawback is that the bootstrap used is not equipped to handle the case with non-stationary common factors.

The current paper can be seen as an attempt to overcome the drawbacks of both these approaches. We begin by developing alternative representations of a cointegrated panel that allows for the possibility of non-stationary common factors. In particular, starting from the triangular representation of the system used by for example Bai *et al.* (2008), we derive a Granger type representation theorem that is similar to the one obtained by Cappuccio and Lubian (1996) in the case of a single time series.

The Granger representation theorem provides not only moving average (MA) and autoregressive moving average (ARMA) representations of the system, but also the conditional ECM representation, which we use as a basis for developing tests for the null hypothesis of no error correction. In particular, paralleling the development of the time series literature in this field, as pioneered by Banerjee *et al.* (1998) and Boswijk (1994), we consider both a t-ratio type test, as well as a Wald type test. Besides eliminating the need for a common factor assumption and a stepwise testing procedure, as shown by Pesavento (2004), these tests are not only more powerful than most residual-based tests around, but are also not worse in terms of size distortions.

It is shown that at the level of the individual unit the asymptotic distribution of the Wald tests is free of nuisance parameters and only depends on the number of non-stationary variables in the system. For the t-ratio an appropriate correction has to be employed to remove the nuisance parameter dependence from the limiting distribution. Nevertheless, because of the common factors, the individual tests are not independent, which of course makes pooling, or cross-sectional averaging, difficult, as it invalidates the use of the conventional limit theory. However, although not analytically tractable, the average still converges to a random variable with a distribution that can be easily simulated, which makes pooling possible in spite of the dependence. We begin by considering the case when the common factors are known, and then we show how the results extend to the case when the factors are approximated by means of cross-sectional averages of the observed data, as suggested by Pesaran (2007).

The rest of this paper is organized as follows. Section 2 presents the model of interest and our version of the Granger representation theorem. Sections 3 and 4 then present the error correction tests and their asymptotic properties, which are verified using both simulated and real data in Sections 5 and 6, respectively. Section 7 concludes.

A word on notation. The symbols $\xrightarrow{w}$ and $\xrightarrow{p}$ will be used to signify weak convergence and convergence in probability, respectively. As usual, $X_T = O_p(T^r)$ will be used to signify that X_T is at most order T^r in probability, while $X_T = o_p(T^r)$ will be used in case X_T is of smaller order in probability than T^r. In the case of a double indexed sequence $X_{N,T}$, $N, T \to \infty$ will be used to signify that the limit has been taken while passing both indices to infinity jointly. For a square matrix A, $rk(A)$, $adj(A)$ and $||A||$ will denote its rank, adjoint and Euclidian norm, respectively. For simplicity, the Brownian motion $B(s)$ defined on the interval $s \in [0,1]$ will be written B, with the measure of integration omitted. We write the integral $\int_0^1 B(s)ds$ as $\int B$ and $\int_0^1 B(s)dB(s)'$ as $\int BdB'$. Finally, $\lfloor x \rfloor$ will be used to denote the integer part of x.

2 Model representation

In this section we discuss the model under consideration, and some alternative representations thereof. We start from the triangular representation for a single unit i, which is the same as the one used by Bai *et al.* (2008). However, these authors focus on how to conduct inference if the variables are in fact long-run related, and do not consider the problem of how to test for cointegration. Moreover, the triangular representation is taken as given, and there is no consideration of other alternatives. Thus, the results reported herein can in many ways be seen as complementary to those reported in Bai *et al.* (2008).

The data generating process has two basic building blocks, a $(r+m)$-dimensional vector of idiosyncratic variables, which is denoted by $Z_{i,t} = (Y_{i,t}', X_{i,t}')'$, where $Y_{i,t}$ is $r \times 1$ while $X_{i,t}$ is $m \times 1$, and a k-dimensional vector of common factors, which is denoted by F_t. The grand vector containing all three variables is denoted $Z_{i,t}^{+} = (Z_{i,t}', F_t')'$, and for later use we will also let $V_{i,t} = (X_{i,t}', F_t')'$ denote the augmented $X_{i,t}$ vector.

The data generating process can be written in the following way

$$
\begin{aligned}
Y_{i,t} - \pi_{1i}' G_t &= b_i' X_{i,t} + \lambda_{1i}' F_t + u_{1i,t}, && (1)\\
\Delta X_{i,t} - \pi_{2i}' g_t &= \lambda_{2i}' \Delta F_t + u_{2i,t}, && (2)\\
\Delta F_t - \pi_3' g_t &= f_t, && (3)
\end{aligned}
$$

where G_t and g_t are vectors of deterministic components such that $g_t = \Delta G_t$ with associated coefficients $\pi_i = \begin{pmatrix} \pi_{1i} & \pi_{2i} & \pi_3 \end{pmatrix}$.

We further assume that the vector $u_{i,t}^{+} = (u_{1i,t}', u_{2i,t}', f_t')'$ is a stationary linear process given by

$$
\begin{aligned}
u_{i,t}^{+} &= \begin{pmatrix} \Gamma_{11i}(L) & \Gamma_{12i}(L) & 0 \\ \Gamma_{21i}(L) & \Gamma_{22i}(L) & 0 \\ 0 & 0 & \Psi(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1i,t} \\ \varepsilon_{2i,t} \\ \eta_t \end{pmatrix} = \begin{pmatrix} \Gamma_i(L) & 0 \\ 0 & \Psi(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{i,t} \\ \eta_t \end{pmatrix} \\
&= \Gamma_i^{+}(L) \varepsilon_{i,t}^{+}, && (4)
\end{aligned}
$$

where $\Psi(L) = I_k - \sum_{j=1}^{\infty} \Gamma_{33j} L^j$ and L is the lag operator. Similarly,

$$\Gamma_i(L) = I_{(r+m)} - \sum_{j=1}^{\infty} \Gamma_{ij} L^j.$$

Equations (1) to (4) constitute the triangular representation of the model. The rest of the assumptions can be summarized in the following way, where $M < \infty$ denotes a generic positive real number.

Assumption 1. *(i)* $\eta_t \sim i.i.d.(0, I_k)$ *with finite fourth moments,* *(ii)* $\sum_{j=0}^{\infty} j \cdot \|\Psi_j\| < M$, *(iii)* $rk(\Psi(1)) = k$.

Assumption 2. *(i)* $\varepsilon_{i,t} \sim i.i.d.(0, \Sigma_i)$ *with finite eighth moments and*

$$\Sigma_i = \begin{pmatrix} \Sigma_{11i} & \Sigma_{12i} \\ \Sigma_{21i} & \Sigma_{22i} \end{pmatrix} = \mathrm{cov}(\varepsilon_{i,t}),$$

(ii) $E(\varepsilon_{i,t}\varepsilon_{j,s}) = 0$ *for all* $i \neq j$ *and* $t \neq s$, *(iii)* $\Gamma_i(L)$ *fulfils the random coefficient and summability conditions of Phillips and Moon (1999, Assumptions 1 and 2),* *(iv)* $rk(\Gamma_i(1)) = r + m$.

Assumption 3. *(i)* $\Lambda_i = (\lambda_{1i}, \lambda_{2i})'$ *is a random matrix such that* $\|\Lambda_i\| < M$, *(ii)* $\overline{\Lambda} = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i \to \mathsf{E}(\Lambda_i) = \Lambda < M$ *as* $N \to \infty$, *(iii)* $rk(\overline{\Lambda}) = k \leq r + m$.

Assumption 4. η_t, $\varepsilon_{i,t}$ *and* Λ_i *are mutually independent.*

Assumptions 1, 2 and 4 imply that for any i, $\varepsilon_{i,t}^{+} \sim i.i.d.(0, \Sigma_i)$ with

$$\Sigma_i^{+} = \begin{pmatrix} \Sigma_i & 0 \\ 0 & I_k \end{pmatrix} = \mathrm{cov}(\varepsilon_{i,t}^{+}).$$

They also imply that $rk(\Gamma_i^{+}(1)) = r + m + k$. Under these assumptions, it is easy to see that the system has r cointegrating relationships $\beta_i' Z_{i,t}^{+}$, where by assumption

$$\beta_i = \begin{pmatrix} I_r & -b_i' & -\lambda_{1i}' \end{pmatrix}'$$

is the cointegrating matrix.

Similar to the time series case considered by Cappuccio and Lubian (1996), given the triangular representation in (1) to (4), we can derive a Granger type representation theorem for a given panel member. This provides us with alternative model representations that are better suited for testing the hypothesis of no cointegration.

Theorem 1. *Given the triangular representation in* (1) *to* (4), $Z_{i,t}^{+}$ *is non-stationary with cointegration rank* r.

(a) The MA representation of $\Delta Z_{i,t}^{+}$ *is*

$$\Delta Z_{i,t}^{+} - (\pi_i^{*})' g_t = C_i(L)\varepsilon_{i,t}^{+}, \tag{5}$$

where $C_i(L)$ *is given in the appendix,* $rk(C_i(1)) = m + k$ *and*

$$\pi_i^{*} = \begin{pmatrix} \pi_{1i} + \pi_{2i} b_i + \pi_3(\lambda_{1i} + \lambda_{2i} b_i) & \pi_{2i} + \pi_3 \lambda_{2i} & \pi_3 \end{pmatrix}.$$

(b) *The ARMA representation of* $Z_{i,t}^{+}$ *is given by*

$$A_i(L)\big(Z_{i,t}^{+} - (\pi_i^*)' g_t\big) = c_i(L)\varepsilon_{i,t}^{+}, \tag{6}$$

where $c_i(L) = |\Gamma_i^{+}(L)|$ *is a scalar lag polynomial, and where the blocks of*

$$A_i(L) = \begin{pmatrix} A_{11i}(L) & A_{12i}(L) & A_{13i}(L) \\ A_{21i}(L) & A_{22i}(L) & A_{23i}(L) \\ 0 & 0 & A_{33i}(L) \end{pmatrix}$$

are given by

$$\begin{aligned}
A_{11i}(L) &= |\Psi(L)||\Gamma_{22i}(L)|adj(\Gamma_{11\cdot 2i}(L)),\\
A_{12i}(L) &= -|\Psi(L)||\Gamma_{22i}(L)|adj(\Gamma_{11\cdot 2i}(L))((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1} + b_i'),\\
A_{13i}(L) &= |\Psi(L)||\Gamma_{22i}(L)|adj(\Gamma_{11\cdot 2i}(L))((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\lambda_{2i}' - \lambda_{1i}'),\\
A_{21i}(L) &= -|\Psi(L)|adj(\Gamma_{22i}(L))\Gamma_{21i}(L)adj(\Gamma_{11\cdot 2i}(L)),\\
A_{22i}(L) &= |\Psi(L)|adj(\Gamma_{22i}(L))\big(\Gamma_{21i}(L)adj(\Gamma_{11\cdot 2i}(L))((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1} + b_i')\\
&+ (1-L)|\Gamma_{11\cdot 2i}(L)|\big),\\
A_{23i}(L) &= -|\Psi(L)|adj(\Gamma_{22i}(L))\big(\Gamma_{21i}(L)adj(\Gamma_{11\cdot 2i}(L))\\
&\times ((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\lambda_{2i}' - \lambda_{1i}') + (1-L)|\Gamma_{11\cdot 2i}(L)|\lambda_{2i}'\big),\\
A_{33i}(L) &= (1-L)|\Gamma_{22i}(L)||\Gamma_{11\cdot 2i}(L)|adj(\Psi(L)),
\end{aligned}$$

with $\Gamma_{11\cdot 2i}(L) = \Gamma_{11i}(L) - \Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\Gamma_{21i}(L)$.

(c) $A_i(1)$ *has reduced rank* r *and can be decomposed as* $A_i(1) = \alpha_i^*\beta_i'$, *where*

$$\alpha_i^* = \begin{pmatrix} |\Psi(1)||\Gamma_{22i}(1)|adj(\Gamma_{11\cdot 2i}(1)) \\ -|\Psi(1)|adj(\Gamma_{22i}(1))\Gamma_{21i}(1)adj(\Gamma_{11\cdot 2i}(1)) \\ 0 \end{pmatrix}.$$

(d) *The vector ECM representation is*

$$A_i^*(L)\big(\Delta Z_{i,t}^{+} - (\pi_i^*)'\Delta g_t\big) = -\alpha_i^*\beta_i'\big(Z_{i,t-1}^{+} - (\pi_i^*)' g_{t-1}\big) + c_i(L)\varepsilon_{i,t}^{+}, \tag{7}$$

where $A_i^*(L) = A_i^{+}(L) + A_i(1)$ *with* $A_i^{+}(L)$ *satisfying* $A_i(L) = A_i(1) + (1-L)A_i^{+}(L)$, $A_i^{+}(L) = \sum_{j=0}^{\infty} A_{ij}^{+} L^J$ *and* $A_{ij}^{+} = -\sum_{l=j+1}^{\infty} A_{il}$.

(e) $\xi_{i,t}' = (Z_{i,t}^{+})'\beta_i$ *has the following representation*

$$\begin{aligned}
\xi_{i,t} &= \beta_i'(\pi_i^*)' G_t + \begin{pmatrix} \Gamma_{11i}(L) & \Gamma_{12i}(L) \end{pmatrix} \varepsilon_{i,t}^{+},\\
\Delta\xi_{i,t} - K_i(L)(\pi_i^*)' g_t &= -\beta_i'\alpha_i^*\big(\xi_{i,t-1} - \beta_i'(\pi_i^*)' g_t\big) + J_i(L)\varepsilon_{i,t}^{+},
\end{aligned}$$

where $K_i(L)$ *and* $J_i(L)$ *can be obtained as in Engle and Granger (1987).*

From the vector ECM representation given in (7) we can obtain the conditional ECM for $Y_{i,t}$ and the marginal ECM for $V_{i,t}$. Towards this end, let $\alpha_i = -A_i(0)^{-1}\alpha_i^*$ and $\tilde{A}_i^*(L) = A_i(0)^{-1}A_i^*(L)$, where $\tilde{A}_i^{**}(L) = \sum_{j=1}^{\infty} \tilde{A}_{ij}^{**} L^j$ with $\tilde{A}_{ij}^{**} = -\tilde{A}_{ij+1}^*$ such that

$$\Delta Z_{i,t}^+ - \tilde{A}_i^*(L)(\pi_i^*)'\Delta g_t = \alpha_i\beta_i'\big(Z_{i,t-1}^+ - (\pi_i^*)'g_{t-1}\big) + \tilde{A}_i^{**}(L)\Delta Z_{i,t-1}^+ + c_i(L)\varepsilon_{i,t}^+.$$

Defining $B_i^* = \big(\Sigma_{12i}\Sigma_{22i}^{-1} + b_i', -\Sigma_{12i}\Sigma_{22i}^{-1}\lambda_{2i}' + \lambda_{1i}'\big)$ and $\kappa_i = (I_r, -B_i^*)$, the conditional ECM for $Y_{i,t}$ is given by

$$\begin{aligned} \Delta Y_{i,t} - \kappa_i\tilde{A}_i^*(L)(\pi_i^*)'\Delta g_t &= B_i^*\Delta V_{i,t} + \kappa_i\alpha_i\beta_i'\big(Z_{i,t-1}^+ - (\pi_i^*)'g_{t-1}\big) + \kappa_i\tilde{A}_i^{**}(L)\Delta Z_{i,t-1}^+ \\ &+ c_i(L)\varepsilon_{1\cdot 2i,t}, \end{aligned} \tag{8}$$

where $\varepsilon_{1\cdot 2i,t} = \varepsilon_{1i,t} - \Sigma_{12i}\Sigma_{22i}^{-1}\varepsilon_{2i,t}$, while the marginal models for $X_{i,t}$ and F_t are

$$\Delta X_{i,t} - \tilde{A}_{2i}^*(L)(\pi_i^*)'\Delta g_t = \alpha_{2i}\beta_i'\big(Z_{i,t-1}^+ - (\pi_i^*)'g_{t-1}\big) + \tilde{A}_{2i}^{**}(L)\Delta Z_{i,t-1}^+ + c_i(L)\varepsilon_{2i,t}^*, \tag{9}$$

$$\Delta F_t - \tilde{A}_{33i}^*(L)\pi_3'\Delta g_t = A_{33i}^{**}(L)\Delta F_{t-1} + c_i(L)\eta_t, \tag{10}$$

where $\tilde{A}_{2i}^*(L)$ and $\tilde{A}_{2i}^{**}(L)$ are the second rows of $\tilde{A}_i^*(L)$ and $\tilde{A}_i^{**}(L)$, respectively, and where $\varepsilon_{2i,t}^* = \varepsilon_{2i,t} + \lambda_{2i}'\eta_t$.

Some remarks can be made here.

Remark 1. What this theorem shows is that alternative representations may lead naturally to alternative approaches to cointegration testing. In particular, while the triangular representation is better suited for developing residual-based tests, the vector ECM, and more precisely its factorization into conditional and marginal models, is more suitable for developing tests based on error correction.

Remark 2. If $\Gamma_i^+(L)$ is a unimodular matrix polynomial, the MA part in the vector ECM in (7) vanishes. Furthermore, if $\Gamma_i^+(L)$ is of order p_i, $A_i(L)$ is of order $q_i \leq (r+m+k-1)p_i$.

Remark 3. The common factor F_t is by assumption strongly exogenous for β_i, see for example Urbain (1992) for weak and strong exogeneity conditions in this class of models. Similarly, $X_{i,t}$ is weakly exogenous for β_i if $\alpha_{2i}^* = 0$, which will be the case when $\Gamma_{21i}(1) = 0$. It is strongly exogenous if in addition $\Gamma_{21i}(L) = 0$. The relevance of the two latter assumptions will be discussed later.

Remark 4. Depending on the specification of the deterministic component g_t, we can distinguish at least five variations of the ECM in (8) to (10). If $g_t = 0$, henceforth referred to as Model 1, then there are no deterministic components present. If $\pi_{1i} = 0$, then $\beta_i'(\pi_i^*)' = 0$ and hence g_t do not appear in the error correction term. If in addition $g_t = (1,t)'$, then a constant should be included, while if $g_t = (1,t,t^2)'$, then a linear trend should also be included.

These specifications are henceforth referred to as Models 2 and 3, respectively. Moreover, if $\pi_{1i} \neq 0$, we have a constant restricted to the error correction term if $g_t = 1$, henceforth referred to as Model 4, or an unrestricted constant and a linear trend in the error correction term if $g_t = (1, t)'$, henceforth referred to as Model 5. Although higher order trend terms are certainly possible, such models are rarely used in practice, and we therefore restrict our attention to these five.

3 Individual tests for no error correction

In this section we show how the conditional ECM in (8) can be used as a basis for constructing cointegration tests. In particular, we propose two test statistics that are designed to test the null hypothesis that unit i is not error correcting versus the alternative that it is error correcting. We begin by considering the baseline case with known factors, and then we show how the testing can be carried out in the more realistic case when F_t is no longer observed.

3.1 Observed factors

Assumptions 1 to 4 are quite relaxed in the sense that even at the level of the individual unit, the models they imply are multivariate, which makes a full-blown system approach necessary. However, the purpose of this section is not to devise the most general test possible, but rather to derive tests that are simple, and easy to implement. This requires more assumptions.

Assumption 5. *(i) $r = 1$, (ii) $c_i(L) = c_i$ for some constant $c_i < M$, (iii) $X_{i,t}$ is weakly exogenous for α_{1i} and β_i.*

Remark 5. Assumption 5 implies that the r-dimensional conditional model in (8) can be written as a well-specified single equation, with no serial correlation and with the scalar coefficient α_{1i} measuring the extent of the error correcting behavior in $Y_{i,t}$.

Under Assumption 5, and omitting any deterministic component for now, the conditional ECM in (8) reduces to

$$\Delta Y_{i,t} = \alpha_{1i}\beta_i' Z_{i,t-1}^{+} + B_{11i}(L)\Delta Y_{i,t-1} + B_{12i}(L)\Delta X_{i,t} + B_{13i}(L)\Delta F_t + \varepsilon_{1\cdot 2i,t}, \tag{11}$$

while the marginal models for $X_{i,t}$ and F_t become

$$\begin{aligned} \Delta X_{i,t} &= B_{21i}(L)\Delta Y_{i,t-1} + B_{22i}(L)\Delta X_{i,t-1} + B_{23i}(L)\Delta F_{t-1} + \varepsilon_{2i,t}^{*}, && (12)\\ \Delta F_t &= B_{33i}(L)\Delta F_{t-1} + \eta_t, && (13) \end{aligned}$$

where the lag polynomials $B_{jli}(L)$ are obtained by simply collecting the appropriate terms from (8) to (10).

Assumptions 1 to 5 ensure that the following functional central limit theorem holds as $T \to \infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} \begin{pmatrix} \varepsilon_{1\cdot 2i,t} \\ \varepsilon_{2i,t}^{*} \\ \eta_t \end{pmatrix} \xrightarrow{w} B_i,$$

where $s \in [0,1]$ and $B_i = (B_{1i}, B_{2i}', B_3')'$ is a $(1+m+k)$-dimensional vector Brownian motion, which can be partitioned as $B_i = (B_{1i}, B_{2\cdot i}')'$ with $B_{2\cdot i} = (B_{2i}', B_3')'$ having dimension $m+k$. The covariance matrix of B_i is given by

$$\Xi_i = \begin{pmatrix} \sigma_i^2 & 0 & 0 \\ 0 & \Sigma_{22i} + \lambda_{2i}'\lambda_{2i} & \lambda_{2i}' \\ 0 & \lambda_{2i} & I_k \end{pmatrix} = \mathrm{cov}(B_i),$$

where $\sigma_i^2 = \Sigma_{11i} - \Sigma_{12i}\Sigma_{22i}^{-1}\Sigma_{21i}$. Thus, $B_i = \Xi_i^{\frac{1}{2}} W_i$, where W_i is a $(1+m+k)$-dimensional standard Brownian motion that is partitioned conformably with B_i. Furthermore, the long-run covariance matrix of $Z_{i,t}^{+}$ is given by

$$\Omega_i = \tilde{B}_i(1)\Xi_i\tilde{B}_i(L)' = \Omega_i^{\frac{1}{2}}(\Omega_i^{\frac{1}{2}})',$$

where the lag polynomial $\tilde{B}_i(L)$ is obtained from collecting the appropriate terms from (11) to (13) and $\Omega_i^{\frac{1}{2}} = \tilde{B}_i(1)\Xi_i^{\frac{1}{2}}$.

For later reference it is useful to consider the continuous time regression of W_{1i}, the first element of W_i, onto some vector X_i,

$$W_{1i} = P_i(X_i)'X_i + Q_X W_{1i},$$

where

$$P_i(X_i) = \left(\int X_i X_i'\right)^{-1} \int X_i W_{1i} = V(X_i)p_i(X_i) \tag{14}$$

is the ordinary least squares (OLS) projection with $Q_X W_{1i}$ being the associated projection error. For example, if $X_i = 1$, then $P_i(X_i) = \int W_{1i}$ in which case $Q_1 W_{1i} = W_{1i} - \int W_{1i}$ is the demeaned version of W_{1i}.

As (11) makes clear, as long as F_t is observed, the problem of testing the null of no error correction is equivalent to testing

$$H_{0i} : \alpha_{1i} = 0$$

against

$$H_{1i} : \alpha_{1i} < 0.$$

The problem is that, unless one resorts to nonlinear techniques, this parameter is not easily estimated. One way to get around this is to assume that β_i is known, and to estimate α_{1i} using OLS. However, as shown by Boswijk (1994) and Zivot (2000), apart from the obvious drawback that β_i is almost never known in practice, tests based on a prespecified β_i are generally not similar and depend on nuisance parameters, even asymptotically.

As an alternative approach, note that (11) can be reparameterized as

$$\begin{aligned}\Delta Y_{i,t} &= \alpha_{1i}Y_{i,t-1} + \gamma_{1i}'X_{i,t-1} + \gamma_{2i}'F_{t-1} + B_{11i}(L)\Delta Y_{i,t-1} + B_{12i}(L)\Delta X_{i,t} \\ &+ B_{13i}(L)\Delta F_t + \varepsilon_{1\cdot 2i,t}, \end{aligned} \tag{15}$$

where $\gamma_{1i}' = -\alpha_{1i}b_i'$ and $\gamma_{2i}' = -\alpha_{1i}\lambda_{1i}'$. The advantage of rewriting (11) in this way is that because γ_{1i} and γ_{2i} are unrestricted, the cointegrating vector is implicitly estimated under the alternative hypothesis. Hence, as long as we are not interested in β_i, all the parameters of (15) can be consistently estimated by simple OLS, which in turn suggests the OLS estimator of α_{1i} as a natural candidate for constructing asymptotically similar tests of the null hypothesis of no error correction. In this section we propose two such tests, whose construction is described next.

One obvious candidate is the t-test. Suppose that the lag polynomial $B_{1ji}(L)$ is of order q_i, and let

$$W_{i,t} = \left(\Delta Y_{i,t-1}, \ldots, \Delta Y_{i,t-q_i}, \Delta X_{i,t}', \ldots, \Delta X_{i,t-q_i}', \Delta F_t', \ldots, \Delta F_{t-q_i}'\right)'$$

denote the vector of stationary, first-differenced, regressors, while $V_{i,t}$ again denotes the vector of weakly exogenous non-stationary, level, variables, then (15) can be written as

$$\begin{aligned}\Delta Y_{i,t} &= \alpha_{1i}Y_{i,t-1} + \gamma_i'V_{i,t-1} + \Pi_i'W_{i,t} + \varepsilon_{1\cdot 2i,t} \\ &= \alpha_{1i}Y_{i,t-1} + \Phi_i'S_{i,t} + \varepsilon_{1\cdot 2i,t}, \end{aligned} \tag{16}$$

where $\Phi_i = (\gamma_i', \Pi_i')'$, $S_{i,t} = (V_{i,t-1}', W_{i,t}')'$, $\gamma_i = (\gamma_{1i}', \gamma_{2i}')'$ and Π_i is the vector stacking the coefficient vectors of the lag polynomials $B_{11i}(L)$, $B_{12i}(L)$ and $B_{13i}(L)$. This equation can in turn be written as

$$\Delta(Q_S Y_{i,t}) = \alpha_{1i}(Q_S Y_{i,t-1}) + Q_S\varepsilon_{1\cdot 2i,t},$$

where again Q_S is the OLS projection error operator, with

$$Q_S Y_{i,t} = Y_{i,t} - \sum_{t=2}^{T} Y_{i,t-1}S_{i,t}' \left(\sum_{t=2}^{T} S_{i,t}S_{i,t}'\right)^{-1} S_{i,t}$$

being the residual from projecting $Y_{i,t}$ onto $S_{i,t}$.

In this notation, the OLS estimator of α_{1i} is given by

$$\hat{\alpha}_{1i} = \left(\sum_{t=2}^{T}(Q_S Y_{i,t-1})^2\right)^{-1} \sum_{t=2}^{T} Q_S Y_{i,t-1}\Delta(Q_S Y_{i,t}),$$

whose estimated variance is given by

$$\mathrm{var}(\hat{\alpha}_{1i}) = \hat{\sigma}_i^2 \left(\sum_{t=2}^{T} (Q_S Y_{i,t-1})^2 \right)^{-1},$$

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=2}^{T}(\Delta(Q_S Y_{i,t}) - \hat{\alpha}_{1i}(Q_S Y_{i,t-1}))^2$. The t-statistic for testing H_{0i} can now be written as

$$\tau_{\hat{\alpha}_{1i}} = \frac{\hat{\alpha}_{1i}}{\sqrt{\mathrm{var}(\hat{\alpha}_{1i})}}.$$

Another possibility is to follow Boswijk (1994), and to use a Wald statistic to test if α_{1i} and γ_i are jointly zero. In so doing, note that (16) can be rewritten as

$$\Delta Y_{i,t} = \delta_{1i}' Z_{i,t-1}^{+} + \Pi_i' W_{i,t} + \varepsilon_{1\cdot 2i,t},$$

where $\delta_{1i} = (\alpha_{1i}, \gamma_i')'$, or in terms of projection residuals,

$$\Delta(Q_W Y_{i,t}) = \delta_{1i}'(Q_W Z_{i,t-1}^{+}) + Q_W \varepsilon_{1\cdot 2i,t}.$$

The Wald statistic for testing the restriction that $\delta_{1i} = 0$ is given by

$$w_{\hat{\delta}_{1i}} = \hat{\delta}_{1i}' \big(\mathrm{var}(\hat{\delta}_{1i})\big)^{-1} \hat{\delta}_{1i},$$

where

$$\hat{\delta}_{1i} = \left(\sum_{t=2}^{T} Q_W Z_{i,t-1}^{+} (Q_W Z_{i,t-1}^{+})' \right)^{-1} \sum_{t=2}^{T} Q_W Z_{i,t-1}^{+} \Delta(Q_W Y_{i,t})$$

is the OLS estimator of δ_{1i}, and

$$\mathrm{var}(\hat{\delta}_{1i}) = \hat{\sigma}_i^2 \left(\sum_{t=2}^{T} Q_W Z_{i,t-1}^{+} (Q_W Z_{i,t-1}^{+})' \right)^{-1}$$

is the associated variance.

The t-statistic $\tau_{\hat{\alpha}_{1i}}$ and the Wald statistic $w_{\hat{\delta}_{1i}}$ are the two test statistics considered in this paper. Their limiting distributions under the no error correction null are given in the following theorem.

Theorem 2. *Under H_{0i} and Assumptions 1 to 5, as $T \to \infty$*

(a) $w_{\hat{\delta}_{1i}} \xrightarrow{w} D_{i,w} = p_i(W_i)' P_i(W_i)$,

(b) $\tau_{\hat{\alpha}_{1i}} \xrightarrow{w} D_{i,\tau}^{\Omega} = \dfrac{d_i}{\sqrt{D_i}}$,

where $p_i(\cdot)$ *and* $P_i(\cdot)$ *are defined in* (14),

$$\begin{aligned}
D_i &= \sigma_i^2 \omega_{11\cdot 2i}^{-2} V(U_i) \\
&+ \sigma_i^2 \omega_{11\cdot 2i}^{-2} \omega_{11i} V(U_i) \left(\rho_i' (\Omega_{22i}')^{-1} P_i(W_{2\cdot i}) + V(W_{2\cdot i}) p_i(W_{2\cdot i})' \Omega_{22i}^{-1} \rho_i \right) \\
&+ \sigma_i^2 \omega_{11\cdot 2i}^{-2} \omega_{11i}^2 \rho_i' (\Omega_{22i}')^{-1} \left(V(W_{2\cdot i}) + P_i(W_{2\cdot i}) V(U_i) V(W_{2\cdot i}) p_i(W_{2\cdot i})' \right) \Omega_{22i}^{-1} \rho_i, \\
d_i &= \sigma_i \omega_{11\cdot 2i}^{-1} P_i(U_i) + \omega_{11\cdot 2i}^{-1} \omega_{11i} \rho_i' (\Omega_{22i}')^{-1} P_i(W_{2\cdot i})(P_i(U_i) - 1),
\end{aligned}$$

with $\omega_{11\cdot 2i}$, ω_{11i}, Ω_{22i} *and* ρ_i *depending on the parameters of* $\Omega_i^{\frac{1}{2}}$, *as defined in the appendix.*

The asymptotic distribution of $\tau_{\hat{\alpha}_{1i}}$ simplifies substantially if $X_{i,t}$ is strongly exogenous.

Assumption 6. $X_{i,t}$ *is strongly exogenous for* α_{1i} *and* β_i.

This is shown in the following corollary.

Corollary 1. *Under Assumption 6 and the conditions of Theorem 2, as* $T \to \infty$,

$$\tau_{\hat{\alpha}_{1i}} \xrightarrow{w} D_{i,\tau} = \frac{P_i(U_i)}{\sqrt{V(U_i)}}.$$

Theorem 2 shows that the distribution of $w_{\hat{\delta}_{1i}}$ as $T \to \infty$ is nuisance parameter free and only depends on $m + k$, the number of non-stationary exogenous variables in the system. By contrast, the distribution of $\tau_{\hat{\alpha}_{1i}}$ depends on several nuisance parameters, and although these vanish under Assumption 6, strong exogeneity is quite restrictive. Fortunately, as Zivot (2000) points out relying on results obtained by for example Saikkonen (1991), there is a simple modification available that eliminates the nuisance parameters that are there under Assumption 5 (iii). The idea is to model these parameters by making the lag polynomial $B_{12i}(L)$ double-sided, as in

$$B_{12i}(L + L^{-1}) = \sum_{j=-\infty}^{\infty} B_{12ij} L^j,$$

where L^{-1} is the lead operator, which in in turn requires augmenting (16) not only by the lags, but also by the leads of $\Delta X_{i,t}$. If the number of leads is large enough, then the asymptotic distribution of the resulting test statistic is given in Corollary 1.

In this sense, the results in Theorem 1 are basically the same as those provided by Banerjee *et al.* (1998) and Boswijk (1994) for the pure time series case. The proof is therefore very similar. The difference lies with the presence of F_t, which has two effects. One is that the number of unit roots increases from $1 + m$ to $1 + m + k$, which is reflected through W_3 in the asymptotic distribution of the test. The second effect is that the test statistics across units are no longer independent of each other, although the degree of the dependence between all pairs of units is the same.

In the presence of nonzero deterministic constant and trend terms, as in Models 2 to 5, the above theorem needs to be modified in order to obtain similar tests. This requires replacing U_i in (a) and W_i in (b) by their appropriately detrended counterparts. Specifically, U_i and W_i should be demeaned in Model 1, and demeaned and detrended in Model 2. The t-test cannot be used in Models 4 and 5, and so for these models there is only the Wald test. In Model 4, W_i is replaced by $(W_i', 1)'$, while in Model 5, W_i is replaced by $(W_i', 1, s)'$, where s is the limiting trend function, see Boswijk (1994).

Furthermore, under the alternative hypothesis of cointegration, $\tau_{\hat{\alpha}_{1i}} \to -\infty$ whereas $w_{\hat{\delta}_{1i}} \to \infty$ as $T \to \infty$, suggesting that the tests are consistent. A proof of this is provided by Boswijk (1994).

3.2 Unobserved factors

So far we have assumed F_t to be observed, an assumption which is generally not true. To account for this, in a recent unit root paper Bai and Ng (2004) propose using the method of principal components to estimate F_t, and then to use this estimate in place of F_t in the subsequent analysis. This approach has proven very fruitful, and has also been extended to the case of cointegration, see for example Bai *et al.* (2008), Banerjee and Carrio-i-Silvestre (2007), Gengenbach *et al.* (2006) and Westerlund (2008). The problem with this approach is that, regardless of whether one considers unit roots or cointegration, the analysis must be carried out in steps, which means that the estimation error from one step is imported in subsequent steps.

As a response to this, Pesaran (2007) proposes a joint approach, which is based on using cross-sectional averages of the observed variables as proxies for the unobserved common factors. Apart from the advantage that it eliminates the need for a two-step estimation procedure, this approach fits very well with the parametric flavor of our conditional ECM, and it will therefore be used in this paper.

Part (b) of Theorem 1 implies that $Z_{i,t}$ can be written as

$$Z_{i,t} = \Lambda_i F_t + E_{i,t},$$

where Λ_i is the $(1+m) \times k$ matrix of factor loadings, and where $E_{i,t}$ is a vector representing the idiosyncratic component of $Z_{i,t}$. Denoting by $\overline{Z}_t$, $\overline{\Lambda}$ and $\overline{E}_t$ the cross-sectional averages of $Z_{i,t}$, Λ_i and $E_{i,t}$, respectively, it is clear that

$$\overline{Z}_t = \overline{\Lambda} F_t + \overline{E}_t,$$

which, via Assumption 3 (iii) and the fact that $E_{i,t}$ is cross-sectionally independent, suggests that F_t can be written as

$$F_t = (\overline{\Lambda}'\overline{\Lambda})^{-1}\overline{\Lambda}'\overline{Z}_t + (\overline{\Lambda}'\overline{\Lambda})^{-1}\overline{\Lambda}'\overline{E}_t = (\overline{\Lambda}'\overline{\Lambda})^{-1}\overline{\Lambda}'\overline{Z}_t + O_p\left(\frac{1}{\sqrt{N}}\right).$$

The implication is that the common factors can be approximated by the cross-sectional averages $\overline{Z}_t$, and that the resulting approximation error should become negligible as $N \to \infty$. Following this argument, we propose using $\overline{Z}_t$ to approximate F_t. In so doing, it is convenient to let $\tilde{W}_{i,t}$, $\tilde{V}_{i,t}$ and $\tilde{Z}^+_{i,t}$ denote $W_{i,t}$, $V_{i,t}$ and $Z^+_{i,t}$, respectively, with $\overline{Z}_t$ in place of F_t. Starting with (16) the approximate test regression can now be written as

$$\Delta Y_{i,t} = \alpha_{1i} Y_{i,t-1} + \Phi_i' \tilde{S}_{i,t} + \tilde{\varepsilon}_{1\cdot 2i,t},$$

or equivalently,

$$\Delta(Q_{\tilde{S}} Y_{i,t}) = \alpha_{1i}(Q_{\tilde{S}} Y_{i,t-1}) + Q_{\tilde{S}} \tilde{\varepsilon}_{1\cdot 2i,t},$$

where the error $\tilde{\varepsilon}_{1\cdot 2i,t}$ depends on the accuracy of the approximation. Nevertheless, by regressing $\Delta(Q_{\tilde{S}} Y_{i,t})$ on $Q_{\tilde{S}} Y_{i,t-1}$, we obtain another OLS estimator of α_{1i}, which we will henceforth denote by $\tilde{\alpha}_{1i}$. The associated t-statistic of H_{0i} can be written in an obvious notation as

$$\tau_{\tilde{\alpha}_{1i}} = \frac{\tilde{\alpha}_{1i}}{\sqrt{\mathrm{var}(\tilde{\alpha}_{1i})}},$$

while the Wald statistic can be written as

$$w_{\tilde{\delta}_{1i}} = (\tilde{\delta}_{1i})' \big(\mathrm{var}(\tilde{\delta}_{1i})\big)^{-1} \tilde{\delta}_{1i},$$

where $\tilde{\delta}_{1i}$ and $\mathrm{var}(\tilde{\delta}_{1i})$ are defined just as in Section 3.1 but with $Q_{\tilde{W}}$ in place of Q_W.

Theorem 3 provides the limiting null distributions of these test statistics.

Theorem 3. *Under the conditions of Theorem 2, as $N, T \to \infty$,*

(a) $w_{\tilde{\delta}_{1i}} \xrightarrow{w} D_{i,w}$,

(b) $\tau_{\tilde{\alpha}_{1i}} \xrightarrow{w} \tilde{D}^{\Omega}_{i,\tau} = \dfrac{\tilde{d}_i}{\sqrt{\tilde{D}_i}}$,

where $\tilde{d}_i$ and $\tilde{D}_i$ are defined analogously to d_i and D_i but depending on the parameters of $\tilde{\Omega}_i^{\frac{1}{2}}$, as defined in the appendix.

Theorem 3 shows that the asymptotic distributions of $w_{\tilde{\delta}_{1i}}$ is the same as that of $w_{\hat{\delta}_{1i}}$ provided in Theorem 2, which is based on observed factors. The limiting distribution of $\tau_{\tilde{\alpha}_{1i}}$ is similar to that of $\tau_{\hat{\alpha}_{1i}}$ but depending on different nuisance parameter due to the approximation of F_t by $\overline{Z}_t$. The difference is that Theorem 2 only requires that $T \to \infty$. If F_t is not observed, we require $N \to \infty$ as well to ensure that $\overline{Z}_t$ provides a sufficiently good approximation for F_t.

Similarly to the case of observed factors, if $X_{i,t}$ is strongly exogenous the asymptotic distribution of $\tau_{\tilde{\alpha}_{1i}}$ simplifies and is the same as that of $\tau_{\hat{\alpha}_{1i}}$. This is shown in Corollary 2.

Corollary 2. *Under Assumption 6 and the conditions of Theorem 3, as $N, T \to \infty$,*

$$\tau_{\hat{\alpha}_{1i}} \xrightarrow{w} D_{i,\tau}.$$

3.3 Critical values

As in the simple case with cross-sectionally independent units, our tests are one-sided. The t-test is left-tailed, while the Wald test is right-tailed. The difference is that in our case the asymptotic test distribution, and hence also the simulation of the critical values, is complicated by the dependence across i. However, conditional on W_3, the Brownian motion associated with F_t, the random variables $D_{1,w}, \ldots, D_{N,w}$, are identically and independently distributed for all values of N. We say that $D_{1,w}, \ldots, D_{N,w}$ form an exchangeable sequence, similar to for example Pesaran (2007) and Gregoir (2005). Thus, since $D_{i,w}$ is the same for all N, we can just as well set $N = 1$ in the simulations, a finding also confirmed by our results. The same argument applies to $D_{i,\tau}$. However, this is only valid for the limiting distribution of the t-test under strong exogeneity of $X_{i,t}$, or if an appropriate correction is employed to remove the nuisance parameter dependence. Otherwise, the individual test statistics are not identically distributed across i.

The simulated critical values at the 1%, 5% and 10% significance levels are reported in Table 1 for the t-test, and in Table 2 for the Wald test. These are based on making 1,000,000 draws from the limiting test distributions, with normal random walks of length $T = 1,000$. The results are reported for all five deterministic model specifications, and for $m = 1, \ldots, 5$.

4 Panel tests for no error correction

In this section we build on the results of Section 3, and show how these can be used to construct pooled tests for the null of no error correction at the overall panel level. As an example, we will consider the t-statistic in the most simple case with known factors.

4.1 The tests

There are many ways in which one can combine a set of individual test statistics into a pooled test. The by far most common way is to follow Im *et al.* (2003) and to take the average, which for the t-statistic in case of known factors amounts to computing

$$\overline{\tau}_{\hat{\alpha}_1} = \frac{1}{N} \sum_{i=1}^{N} \tau_{\hat{\alpha}_{1i}}.$$

This is a test of the null of no error correction against the alternative that there is a non-vanishing fraction of error correcting units. Formally, the null and alternative hypotheses are formulated as

$$H_0 : \alpha_{1i} = 0 \text{ for all } i$$

against

$$H_1 : \alpha_{1i} < 0 \text{ for } i = 1, \ldots, N_1 \text{ with } \frac{N_1}{N} \to \delta > 0$$

as $N_1, N \to \infty$. However, due to the dependence across i, in our case it is not possible to follow the usual practice in applying a central limit theorem to obtain a normal distribution for $\sqrt{N}$ times $\overline{\tau}_{\hat{\alpha}_1}$.

One possibility is to look directly at the average. Following similar arguments as Pesaran (2007), because $D_{1,\tau}, \ldots, D_{N,\tau}$ are identically and independently distributed given W_3, a law of large numbers applies to the conditional average of these random variables. That is, we have that as $N \to \infty$

$$\overline{D}_\tau = \frac{1}{N} \sum_{i=1}^{N} D_{i,\tau} \xrightarrow{p} \mathsf{E}(D_\tau | W_3),$$

where the i index in the expectation has been suppressed because all $D_{i,\tau}$ have the same conditional expectation. Thus, unconditionally the average converges to some random distribution. However, unless $\tau_{\hat{\alpha}_{1i}}$ has finite moments for all N and T, this distribution is not necessarily the same as the one that applies to $\overline{\tau}_{\hat{\alpha}_1}$.

In order to get around this technical difficulty, we follow Pesaran (2007) and base our pooled test on a truncated version of $\tau_{\hat{\alpha}_{1i}}$. Because this test has finite moments by construction, the associated cross-sectional average converges to the same asymptotic distribution as $\overline{D}_\tau$.

The truncated statistic is defined as

$$\tau^*_{\hat{\alpha}_{1i}} = \begin{cases} K_l & \text{if} \quad \tau_{\hat{\alpha}_{1i}} \leq K_l \\ \tau_{\hat{\alpha}_{1i}} & \text{if} \quad K_l < \tau_{\hat{\alpha}_{1i}} < K_u \\ K_u & \text{if} \quad \tau_{\hat{\alpha}_{1i}} \geq K_u \end{cases},$$

where the thresholds K_l and K_u are such that the probability of observing $\tau_{\hat{\alpha}_{1i}} \leq K_l$ and $\tau_{\hat{\alpha}_{1i}} \geq K_u$ is sufficiently small. In particular, by using the normal approximation of $\tau_{\hat{\alpha}_{1i}}$, $K_l = \mathsf{E}(D_\tau) - \Phi^{-1}\big(1 - \frac{\varepsilon}{2}\big)\sqrt{\text{var}(D_\tau)}$ and $K_u = \mathsf{E}(D_\tau) + \Phi^{-1}\big(1 - \frac{\varepsilon}{2}\big)\sqrt{\text{var}(D_\tau)}$, where $\varepsilon > 0$ is a small number, while Φ is the standard normal cumulative distribution function.

The corresponding truncated version of $\overline{\tau}_{\hat{\alpha}_1}$ is given by

$$\overline{\tau}^*_{\hat{\alpha}_1} = \frac{1}{N} \sum_{i=1}^{N} \tau^*_{\hat{\alpha}_{1i}}.$$

Making use of Theorem 2, it is not difficult to see that as $T \to \infty$

$$\overline{\tau}^*_{\hat{\alpha}_1} \xrightarrow{w} \overline{D}^*_\tau = \frac{1}{N} \sum_{i=1}^{N} D^*_{i,\tau},$$

where

$$D^*_{i,\tau} = \begin{cases} K_l & \text{if} \quad D_{i,\tau} \leq K_l \\ D_{i,\tau} & \text{if} \quad K_l < D_{i,\tau} < K_u \\ K_u & \text{if} \quad D_{i,\tau} \geq K_u \end{cases}.$$

But all moments of $D^*_{i,\tau}$ exist, so by conditioning on W_3, as $N \to \infty$

$$\overline{D}^*_\tau \xrightarrow{p} \mathsf{E}(D^*_\tau | W_3),$$

where

$$\begin{aligned}\mathsf{E}(D_\tau^*|W_3) &= K_l \cdot \mathrm{Prob}(D_\tau \leq K_l|W_3) + K_u \cdot \mathrm{Prob}(D_\tau \geq K_u|W_3) \\ &+ \mathsf{E}(D_\tau|W_3, K_l < D_\tau < K_u) \to \mathsf{E}(D_\tau|W_3)\end{aligned}$$

as K_l, $K_u \to \infty$, and so we get the same result as for $\overline{D}_\tau$. This suggests that $\overline{\tau}^*_{\hat{\alpha}_1}$ can be used for the test of H_0 versus H_1. Another possibility is to use $\overline{w}^*_{\hat{\delta}_1}$, the average of the truncated Wald test statistics.

4.2 Critical values

The above results show that if K_l, $K_u \to \infty$, $\overline{\tau}^*_{\hat{\alpha}_1}$ converges to a distribution that only depends on number of non-stationary variables in the system. With K_l and K_u finite, however, then there is not just this dependence, but also a dependence on the specific threshold values. Similarly, if N is finite, then there is also a dependence on the size of the cross-section. The generation of the critical values has to account for all these dependencies.

We begin by simulating values of $\mathsf{E}(D_\tau)$ and $\mathrm{var}(D_\tau)$ for all five deterministic model specifications, and for $m = 1, \ldots, 5$. These are needed in order to compute K_l and K_u. Just as in Section 3.3 we make 1,000,000 draws from the limiting test distribution, with normal random walks of length $T = 1,000$. The results for the t-test are reported in Table 1, while the results for the Wald test are reported in Table 2.

The next step is to simulate N-tuples $D^*_{1,\tau}, \ldots, D^*_{N,\tau}$ using $\varepsilon = \frac{1}{10^6}$, and the first-step moments to compute K_l and K_u. The average is then taken, which yields one simulated value of $\overline{D}^*_\tau$. By repeating this exercise 10,000 times, we obtain the simulated distribution of $\overline{D}^*_\tau$. The critical values at the 1%, 5% and 10% levels are reported in Table 3 for the t-test and in Table 4 for the Wald test, in which case $\overline{D}^*_\tau$ is replaced by $\overline{D}^*_w$, the average of the truncated Wald test distributions.

5 Monte Carlo simulations

In this section we report the findings of a small set of simulations. We do not intend to give a comprehensive account of all the merits and drawbacks of the tests, but rather we want to convey a rough idea of their relative performance, also when compared to some of the more conventional tests from the literature.

The data generating conditional ECM is given by

$$\Delta Y_{i,t} = \alpha_1 \left(Y_{i,t-1} - X_{i,t-1} - \iota_2' F_{t-1} \right) + \Delta X_{i,t} + B_{13i} \Delta F_t + \varepsilon_{1\cdot 2i,t},$$

while the marginal models for $X_{i,t}$ and F_t are generated as

$$\begin{aligned}\Delta X_{i,t} &= B_{23i} \Delta F_{t-1} + \varepsilon^*_{2i,t}, \\ \Delta F_t &= \eta_t,\end{aligned}$$

where the elements of B_{23i} and B_{13i} are drawn from $N(1,1)$, while $\iota_2 = (1,1)'$ is a two-dimensional vector of ones. Thus, in this setup $X_{i,t}$ is a scalar, while F_t is two-dimensional. For simplicity, we assume that there are no deterministic components in the data generating process, and that there is a common error correction parameter α_1, which is equal to zero under the null hypothesis, and equal to -0.05 under the alternative.

The results are organized in four parts depending on whether there is any serial correlation present or not. If there is no serial correlation, then $\varepsilon_{1\cdot 2i,t}$, $\varepsilon^*_{2i,t}$ and η_t are drawn from the standard normal distribution, while if there is serial correlation, then one of these errors is specified as a first-order autoregressive (AR) process with standard normal innovations, and a common AR coefficient of magnitude 0.5, while the remaining two errors are again drawn from the standard normal distribution.

All experiments are based on generating $5,000$ panels with N individual and $T+50$ time series observations, where the first 50 observations for each series are discarded in order to attenuate the effect of the initial conditions, which are all set to zero.

For comparison, the error correction tests of Westerlund (2007) are also simulated. Two are based on the group mean, or between, principle and are denoted G_τ and G_ρ, while the corresponding panel, or within, type statistics are denoted P_τ and P_ρ. Analogous to $\overline{\tau}^*_{\hat{\alpha}_1}$, G_τ and P_τ are constructed as t-ratios, while G_ρ and P_ρ are coefficient type statistics.

The problem with these tests is that they are based on assuming cross-sectional independence, as explained earlier, and are therefore not expected to work in a setup as general as this one. Therefore, for better comparability, we follow the suggestion of Gengenbach *et al.* (2006), and run the tests on the defactored data. Specifically, we begin by estimating separately the common component of $X_{i,t}$ and $Y_{i,t}$ using the method of Bai and Ng (2004), which involves applying the principal components method to the variables in their first differences. The estimated common component is then removed, and the defactored data are cumulated back to levels again. The number of factors are determined using the IC_1 information criterion of Bai and Ng (2002) with a maximum of five factors.

For the number of lags and leads to use in the conditional ECM, we used the Schwarz Bayesian information criterion, which facilitates a data dependent choice. Consistent with the results of Ng and Perron (1995), the maximum number of lags and leads is permitted to grow with T at rate $4\left(\frac{T}{100}\right)^{2/9}$. The same rate is used for picking the bandwidth needed for constructing G_ρ and P_ρ. Also, for better comparability across all tests, we do not consider Models 4 and 5 when the deterministic constant and trend terms are restricted to the error correction term. All tests are performed at the 5% significance level, and all powers are adjusted for size.

The results for the case with no deterministic components are reported in Table 5. The first thing to note is the relative performance of the new t-tests, which is very good. This is especially true when the data are serially correlated, in which case there are only one other test with roughly the same performance as ours, G_τ. The overall best performance is obtained by using the individual $\tau_{\hat{\alpha}_{1i}}$ and $\tau_{\tilde{\alpha}_{1i}}$ tests, which seem to maintain the nominal level very well in all cases considered. At the other end of the scale we have the $\overline{w}_{\hat{\delta}_1}$ test, which generally suffers from severe distortions, even if it is based on the true factors.

Pesavento (2004) reports some results for the original Wald test of Boswijk (1994), and find it to be oversized when the serial correlation is of the positive AR type considered here. The overall poor performance of the new Wald tests is therefore not very surprising. On the other hand, unreported results suggest that the relative performance of these tests is much improved if the serial correlation is of the negative MA type, which is also what Pesavento (2004) finds in her simulation study. In any case, the size distortions generally decrease substantially as T increases, which corroborates our asymptotic results.[1]

Among the different versions of the new tests considered, the best size accuracy is not surprisingly obtained by using the true factors. The tests based on using the cross-sectional averages of the observed data as proxies for the factors are, however, almost as accurate, and perform only slightly less well. Thus, the approximation seem to be effective even when N is as small as 10. The defactored versions of the tests of Westerlund (2007) also seem to perform quite well, which is in agreement with consistency of the principal components method, as shown by Bai and Ng (2004). However, although improving in N, we also see that the size accuracy is basically unaffected by T, which is unexpected because theoretically the precision of the principal components estimator should get better as T grows.

Consider next the results reported in Table 5 for the power of the tests, which can be summarized as follows. Firstly, the power increases rapidly as T and N increase, which is presumably a reflection of the consistency of the tests. Secondly, the Westerlund (2007) tests generally suffer from poor power, especially when $\varepsilon_{1\cdot 2i,t}$ is serially correlated, in which case the power is only rarely in excess of the size. The G_ρ test suffers most, and can actually be less powerful than some of the individual tests. Thirdly, as expected, the power of the new tests is generally greatly improved by pooling. Similarly, the tests based on the true factors are generally more powerful than those based on the cross-sectional averages of the observed data.

The results for the models with a constant, and constant and trend reported in Tables 6 and 7, respectively, are very similar to those reported in Table 5. Nevertheless, there are still a few differences that are noteworthy. One difference is the magnitude of the size distortions, which has a slight tendency to increase as more deterministic components are added. Similarly, we see that inclusion of more deterministic components reduces the power

[1]One possibility here is to follow Palm *et al.* (2007), and to use bootstrap methods to eliminate the size distortions of the Wald test.

of the tests, especially for the individual ones. Another difference is that the relative power of the Wald tests is generally much higher in Tables 6 and 7 than in Table 5.

We also examined the effects of a violation of the weak exogeneity assumption. We used the same data generating process as before but this time we allowed the equation for $\Delta X_{i,t}$ to be error correcting. The results, which are not reported but available from the corresponding author upon request, conforms well with our expectations. In particular, while the size of the tests is not effected, the power can be very low in cases when it is mainly $\Delta X_{i,t}$ that is error correcting. Thus, even though the tests continue to perform well in some setups, in general we need the weak exogeneity assumption to ensure that they work properly.[2]

The above results are all based on the truncated panel statistics. We carried out the same simulations for their non-truncated versions, and obtained identical results. In fact, the two types of statistics differ only for very small values of T, and are basically indistinguishable for $T > 20$. Thus, although little is gained in the present case, the truncation of the extreme test statistics seem to pay out when T is very small. This effect is particularly strong when the number parameters of the underlying ECM regressions is large.

6 Empirical Applications

In this section we present two empirical applications of the tests developed in this paper. The first is concerned with the Fisher effect, while the second is concerned with the monetary exchange rate model.

6.1 The Fisher effect

There are very few theoretical economic relationships with as much intuitive appeal as the Fisher effect, which states that a one-time permanent shock in monetary variables has no long-run effect on the real economy. A simple implication of this theory is that changes in inflation should be reflected fully in subsequent movements of the nominal interest rate, thus leaving the real interest rate constant over time. Yet, oddly, for a theory so widely accepted, the postulated long-run relationship between inflation and nominal interest rates has proven extremely difficult to establish empirically. In fact, most studies are unable to reject the null hypothesis of no cointegration between inflation and nominal interest rates.

Westerlund (2008) argues that this lack of empirical support can be partly explained by the poor precision of the routinely applied time series approach, and that the use of panel data can produce more accurate tests. Consistent with this story, drawing upon a panel of 20 OECD countries between the first quarter of 1980 and the fourth quarter of 2004, the author shows that while the null hypothesis of no cointegration cannot be rejected at

[2] Zivot (2000) examines the performance of the time series tests of Banerjee *et al.* (1998) and Boswijk (1994), and reach the same conclusions.

conventional significance levels when using data on individual countries, panel testing leads to a safe rejection. Low power in the tests is therefore one possible explanation for why cointegration has been so difficult to find.

Our findings suggest that there is an alternative interpretation of these results. Namely, that inflation and nominal interest rates are cross-sectionally correlated via the presence of non-stationary common factors, which then invalidates the use of conventional critical values.[3] Thus, according to this view, it is the factors, and not a lack of power, that make the tests unable to reject the no cointegration null at the individual country level.[4]

In this section, we therefore apply our new tests to the same data to reevaluate the cointegration test results reported by Westerlund (2008). In so doing, we will assume that his unit root test results hold, and hence that the rates of inflation and nominal interest are non-stationary. Hence, in this application $Y_{i,t} = i_{i,t}$ and $X_{i,t} = \pi_{i,t}$, where $i_{i,t}$ is the nominal interest rate for country i in quarter t, while $\pi_{i,t}$ is inflation.

The tests are constructed in the same way as in Section 5, using the Schwarz Bayesian information criterion with the same maximum to determine the number of lags and leads. One difference in comparison to the simulations is that the common factors are no longer observed, which means that we cannot evaluate the tests at the true factors. Therefore, as a feasible alternative, in this section we consider replacing the factors by their first differenced and cumulated principal components estimates, which are consistent even if the factors are non-stationary, see Bai and Ng (2004). In agreement with the so-called full Fisher effect, the estimation is carried out while imposing a unit slope coefficient on inflation. That is, the factors are estimated from the real interest rate, $i_{i,t} - \pi_{i,t}$, which is consistent with the idea of the existence of a world real interest rate, see for example Lee (2002).

The principal components method is implemented as described in Section 5, but with the number of factors restricted to two, which ensures that the rank condition in Assumption 3 (iii) is fulfilled. As in the simulations, the defactored versions of the error correction tests of Westerlund (2007) are also considered. We focus on the results for Model 2 with an unrestricted constant, but include the results for Model 3 with both constant and trend for comparison.

The results reported in Table 8 suggest that there is strong evidence against the no cointegration null, even at the individual country level, which goes against the power argument of Westerlund (2008). Indeed, looking at the baseline specification with no trend, we end up rejecting the null for 13 out of the 20 countries when using the $\tau_{\tilde{\alpha}_{1i}}$ test, and for 11 countries

[3] Although the panel tests of Westerlund (2008) are immune to the presence of common factors, his time series tests are not. This means that the two sets of results are not really comparable in the sense that the observed non-rejections at the individual country level could be due to the factors.

[4] One rationale for these factors is that they represent in part oil price shocks and other unanticipated changes in inflation.

when using the $w_{\tilde{\delta}_{1i}}$ test. Similarly, the pooled tests are way out in the critical region and lead to a safe rejection, even at the conservative 1% level. In other words, there is not much evidence against the Fisher effect. This conclusion is not altered by the inclusion of a linear trend.

In fact, the no cointegration null is rejected even when the factors are estimated with the slope on inflation fixed at unity. Specifically, although weaker at the individual level, the evidence at the overall panel level is still strong. Thus, we also have some evidence of the full Fisher effect.

To formally test for the presence of unit roots in the estimated factors, we follow the recommendation of Bai and Ng (2004) and use the augmented Dickey and Fuller (1979) test, ADF henceforth.[5] The estimated first order AR coefficient for the two factors are 0.81 and 0.89, indicating that there is considerable persistency in the factors. This evidence is reinforced by the associated ADF test values, -1.69 and -1.79, respectively, which lead to an acceptance of the unit root null for both factors. Thus, if these factors are to be interpreted as emanating from the world real interest rate, then this rate must be non-stationary.

The lesson we draw from these results is that a failure to reject the null of no cointegration at the individual country level need not be taken as an indication of low power, as the possibility remains that it can be due to the presence of non-stationary common factors.

6.2 The monetary exchange rate model

In this section we take a closer look at the monetary exchange rate model, which postulates a strong link between the nominal exchange rate and a set of monetary fundamentals. The by far most scrutinized proposition being that the nominal exchange rate between the domestic and the foreign reference country, usually the United States, should cointegrate with the relative money supply and relative output of these countries.

However, as with the Fisher effect, despite its strong theoretical appeal, the empirical success of the monetary model has been rather limited, to say the least. Westerlund (2008), Mark and Sul (2001) and Rapach and Wohar (2004) for example argue that this is due to low power. They then proceed to show that the use of panel data leads to a much more favorable picture, with strong evidence of cointegration at the aggregate panel level. Therefore, since the countries appear to be cointegrated, the authors proceed to estimate the cointegration vector.

The problem is that since all variables are measured relative to the United States, this means that the common factors are there by construction. Furthermore, both money supply and output are generally believed to possess unit roots, even for the United States, such that

[5]The test allow for an intercept and the lag orders are determined using the Schwarz Bayesian information criterion.

the common factors must be non-stationary as well. The potential consequences of unattended non-stationary factors on residual-based panel cointegration tests have been studied by Banerjee *et al.* (2004) and Gengenbach *et al.* (2006). The effects may lead to size distortions in small samples or even divergence in large samples. While Mark and Sul (2001) employ a block bootstrap to correct for some weak cross section dependence among the error term. It is not clear whether their test can correct for strong cross sectional dependence induced by non-stationary common factors. Rapach and Wohar (2004) only allow for cross section dependence in form of a common time effect.

In this section we revisit the results of Mark and Sul (2001) and Rapach and Wohar (2004). The data are taken directly from Mark and Sul (2001), and cover 18 countries between the first quarter of 1973 and the first quarter of 1997. Thus, in this application, $Y_{i,t} = e_{i,t}$ and

$$X_{i,t} = \begin{pmatrix} m_t^* - m_{i,t} \\ y_t^* - y_{i,t} \end{pmatrix},$$

where $e_{i,t}$, $m_{i,t}$ and $y_{i,t}$ are the logarithm of the nominal exchange rate, money supply and real income for country i at quarter t, respectively. Asterisks denote the United States.

The average-based tests are computed in the same ways as before, but now we consider two new versions of the factor-based tests. The first is based on using m_t^* and y_t^* as observed factors, which is very interesting in the sense that it provides an example of the scenario considered in Section 3.1. The second version is based on pre-specifying the cointegrating relationship as in Mark and Sul (2001). In particular, it is assumed that the relationship can be written as

$$\beta' Z_{i,t} = \begin{pmatrix} 1, & -1, & 1 \end{pmatrix} \begin{pmatrix} Y_{i,t} \\ X_{i,t} \end{pmatrix} = e_{i,t} - (m_t^* - m_{i,t}) + (y_t^* - y_{i,t}),$$

which imposes monetary neutrality and a unit negative income elasticity.[6] Three factors are estimated from this relationship, which again ensures that Assumption 6 is satisfied. Once again we focus on Model 2 with an unrestricted constant as the deterministic component. For simplicity, in this section we drop the Westerlund (2007) tests.

The results are reported in Table 9. The first thing to note is that for the first 11 countries there is almost no evidence of cointegration at all, except possibly for Belgium, where we count four rejections at the 5% level. The pooled tests are generally much more significant, especially the Wald tests, displaying evidence of cointegration for all five panels. Just as before the results show almost no variation at all depending on whether the trend is included or not.

[6]In order to avoid the problem with nuisance parameter dependency discussed in Section 3.1, the cointegrating relationship is only pre-specified for the purpose of estimating the factors. In other words, although restricted in the factor estimation, in the implementation of the error correction tests the cointegrating relationship is still unrestricted.

These two sets of results suggest that the evidence at the aggregate panel level could very well be due to only a few cointergrated countries. Indeed, a closer look at the different panel members reveals that the significance at the aggregate panel level is mainly due to three individually cointegrated countries, Italy, Spain and Korea. Although these differences could of course also be due to the relatively low power of the individual tests, they nevertheless show that one should take caution in interpreting test results at the aggregate panel level. Indeed, based on the results reported here it seems very hazardous, and borderline erroneous, to treat all five panels as cointegrated, and to proceed with the analysis as if all members are individually cointegrated.

When we compare the results from across the different tests, in agreement with our simulations, we see that the average-based Wald test leads to most rejections. As a final piece of evidence, Table 10 reports some summary statistics for the estimated factors. As in the Fisher application, we see that the estimated AR coefficients are very close to one, indicating the presence of unit roots, which is again supported by the ADF test results.

7 Conclusions

In this paper we consider the issue of testing for cointegration in a panel data model with non-stationary common factors. We begin by showing that the model admits to an ECM representation, a result that is then used for developing two new cointegration tests based on the significance of the error correction term.

It is shown that under the null of no error correction the asymptotic distributions of the tests are free of nuisance parameters, and that they only depend on the number of non-stationary variables in the system. However, the individual tests are not independent along the cross-sectional dimension, which makes pooling difficult. Nonetheless, the cross-sectional averages of these tests are shown to converge to well-defined distributions. These results hold regardless of whether the factors are treated as known or if they are estimated using the averages of the observed data. Some simulation evidence is also provided showing that the tests behave quite well in small samples.

A number of concluding remarks can be made. Firstly, the assumption of weak exogeneity of the regressors in the ECM is crucial for correct interpretation of the tests. This assumption is clearly a weakness in comparison to the residual-based test approach, in which the regressors can be fully endogenous by means of a non-parametric correction. However, it should be pointed out that in principle there is nothing that precludes the use of a similar correction in the current setup. An alternative approach would be to pre-test the validity of the weak exogeneity assumption using panel extensions of the Lagrange multiplier tests proposed by Boswijk and Urbain (1997).

Secondly, the simulations show that the new tests can still be distorted in some cases when the factors are treated as unknown. One possibility towards this end would be to follow Palm *et al.* (2007), and to consider bootstrap versions of our tests, which are expected to have better size properties in small samples.

Finally, a crucial assumption is that of a single cointegrating vector under the alternative. This is obviously an important limitation of our tests that is shared with most existing residual-based tests. When the dimension of the cointegrating space is unknown it is probably best to analyze the data using system-based approaches, see for example Larsson *et al.* (2001).

References

[1] Bai, J. and J. L. Carrion-i-Silvestre (2007). Testing for Panel Cointegration with Unobservable Dynamic Common Factors. Unpublished Manuscript.

[2] Bai, J., C. Kao and S. Ng. (2008). Panel Cointegration with Global Stochastic Trends. Unpublished Manuscript.

[3] Bai, J., and S. Ng. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* **70**, 191–221.

[4] Bai, J., and S. Ng. (2004). A PANIC Attack on Unit Roots and Cointegration. *Econometrica* **72**, 1127–1177.

[5] Banerjee, A. and J. L. Carrion-i-Silvestre (2007). Cointegration in Panel Data with Breaks and Cross-Section Dependence. Economics Working Papers ECO2006/5, European University Institute.

[6] Banerjee, A., M. Marcellino, and C. Osbat (2004). Some Cautions on the Use of Panel Methods for Integrated Time Series of Macroeconomic Data. *Econometrics Journal* **7**, 322–340.

[7] Boswijk, P. H. (1994). Testing for an Unstable Root in Conditional and Structural Error Correction Models. *Journal of Econometrics* **63**, 37–60.

[8] Boswijk, P. H., and J.-P. Urbain (1997). Lagrange Multiplier Tests for Weak Exogeneity: A Synthesis. *Econometric Reviews* **16**, 21–38.

[9] Breitung, J., and M. H. Pesaran (2008). Unit Roots and Cointegration in Panels. In Matyas, L., and P. Sevestre (Eds.) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, 279–322. Kluwer Academic Publishers, Boston.

[10] Cappuccio, N., and D. Lubian (1996). Triangular Representation and Error Correction Mechanism in Cointegrated Systems. *Oxford Bulletin of Economics and Statistics* **58**, 409–415.

[11] Engle, R. F., and C. W. J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation and Testing. *Econometrica* **55**, 251–276.

[12] Gengenbach, C., F. C. Palm and J.-P. Urbain (2006). Panel Cointegration Testing in the Presence of Common Factors. *Oxford Bulletin of Economics and Statistics* **68**, pp. 683-719.

[13] Gregoir, S. (2005). Representaion and Statistical Analysis of Weakly Linearly Exchangable Dynamic Panels. Working Paper, CREST.

[14] Im, K. S., M. Peseran and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115**, 53–74.

[15] Larsson, R., J. Lyhagen and M. Lötgren (2001). Likelihood-Based Cointegration Test in Heterogeneous Panels. *Econometrics Journal* **4**, 109–142.

[16] Lee, J. E. (2002). On the Characterisation of the World Real Interest Rate. *World Economy* **25**, 247–255.

[17] Mark, N. C., and D. Sul (2001). Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Panel. *Journal of International Economics* **53**, pp. 29-52.

[18] Ng, S., and P. Perron (1995). Unit Root Test in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Association* **90**, 268–281.

[19] Palm, F., S. Smeekes and J.-P. Urbain (2007). A Sieve Bootstrap Test for Cointegration in a Conditional Error Correction Model. METEOR Research Memorandum 07054, Maastricht University.

[20] Park, J. Y., and P. C. B. Phillips (1988). Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory* **4**, 468–497.

[21] Pedroni, P. (1999). Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics* **61**, 653–670.

[22] Pedroni, P. (2004). Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis. *Econometric Theory* **20**, 597–625.

[23] Pesaran, M. H. (2007). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.

[24] Pesavento, E. (2004). Analytical Evaluation of the Power of Tests for the Absence of Cointegration. *Journal of Econometrics* **122**, 349–384.

[25] Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.

[26] Rapach, D. E., and M. E. Wohar (2004). Testing the Monetary Model of Exchange Rate Determination: A Closer Look at Panels. *Journal of International Money and Finance* **23**, 867–895.

[27] Saikkonen, P. (1991). Asymptotically Efficient Estimation of Cointegration Regressions. *Econometric Theory* **7**, 1–21.

[28] Urbain, J.-P. (1992). On Weak Exogeneity in Error Correction Models, *Oxford Bulletin of Economics and Statistics* **54**, 187–206.

[29] Westerlund, J. (2007). Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics* **69**, 709–748.

[30] Westerlund, J. (2008). Panel Cointegration Tests of the Fisher Effect. *Journal of Applied Econometrics* **23**, 193–233.

[31] Westerlund, J., and R. Larsson (2008). A Note on the Pooling of Individual PANIC Unit Root Tests. Unpublished manuscript.

[32] Zivot, E. (2000). The Power of Single Equation Tests for Cointegration when the Cointegration Vector is Prespecified. *Econometric Theory* **16**, 407–439.

A Appendix

A.1 Proof of Theorem 1.

Consider (a). From (3) and (4) we have $\Delta F_t - \pi_3' g_t = \Psi(L)\eta_t$. Substituting for ΔF_t in (2) and using (4) we obtain

$$\Delta X_{i,t} - \pi_{2i}^{*}{}' g_t = \Gamma_{21i}(L)\varepsilon_{1i,t} + \Gamma_{22i}(L)\varepsilon_{2i,t} + \lambda_{2i}'\Psi(L)\eta_t. \tag{A1}$$

Taking first differences of (1) and substituting for ΔF_t and $\Delta X_{i,t}$ from (A1) we obtain the following MA representation

$$\begin{aligned}\Delta Y_{i,t} - \pi_{1i}^{*}{}' g_t &= ((1-L)\Gamma_{11i} b_i' \Gamma_{21i}(L))\varepsilon_{1i,t} \\ &\quad + ((1-L)\Gamma_{21i} + b_i'\Gamma_{22i}(L))\varepsilon_{2i,t} + (\lambda_{1i}' + b_i'\lambda_{2i}')\Psi(L)\eta_t. \end{aligned} \tag{A2}$$

Combining the results above we find

$$C_i(L) = \begin{pmatrix} (1-L)\Gamma_{11i}(L) + b_i'\Gamma_{21i}(L) & (1-L)\Gamma_{12i}(L) + b_i'\Gamma_{22i}(L) & (\lambda_{1i}' + b_i'\lambda_{2i}')\Psi(L) \\ \Gamma_{21i}(L) & \Gamma_{22i}(L) & \lambda_{2i}'\Psi(L) \\ 0 & 0 & \Psi(L) \end{pmatrix}.$$

Furthermore,

$$C_i(1) = \begin{pmatrix} b_i' & (\lambda_{1i}' + b_i'\lambda_{2i}') \\ I_m & \lambda_{2i}' \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \Gamma_{21i}(1) & \Gamma_{22i}(1) & 0 \\ 0 & 0 & \Psi(1) \end{pmatrix}$$

such that $C_i(1)$ has rank $m+k$. This establishes part (a) of the theorem.

Next, consider (b). Partition $C_i(L)$ such that the diagonal blocks $C_{11i}(L)$ and $C_{22i}(L)$ are of dimension $r \times r$ and $(m+k) \times (m+k)$, respectively. Since

$$C_{22i}(L) = \begin{pmatrix} \Gamma_{22i}(L) & \lambda_{2i}'\Psi(L) \\ 0 & \Psi(L) \end{pmatrix}$$

is invertible, we can decompose $C_i(L)$ as

$$C_i(L) = \begin{pmatrix} I_r & C_{12i}(L)C_{22i}(L)^{-1} \\ 0 & I_{(m+k)} \end{pmatrix} \begin{pmatrix} C_{11\cdot 2i}(L) & 0 \\ C_{21i}(L) & C_{22i}(L) \end{pmatrix}.$$

As $C_{11\cdot 2i} = (1-L)\Gamma_{11\cdot 2i}(L)$ we can further write

$$\begin{aligned} C_i(L) &= \begin{pmatrix} I_r & C_{12i}(L)C_{22i}(L)^{-1} \\ 0 & I_{(m+k)} \end{pmatrix} \begin{pmatrix} (1-L)I_r & 0 \\ 0 & I_{(m+k)} \end{pmatrix} \begin{pmatrix} \Gamma_{11\cdot 2i}(L) & 0 \\ C_{21i}(L) & C_{22i}(L) \end{pmatrix} \\ &= U_i(L)^{-1} M(L) V_i(L)^{-1}, \end{aligned} \tag{A3}$$

where the lag polynomials

$$V_i(L) = \begin{pmatrix} \Gamma_{11\cdot 2i}(L)^{-1} & 0 & 0 \\ -\Gamma_{22i}(L)^{-1}\Gamma_{21i}(L)\Gamma_{11\cdot 2i}(L)^{-1} & \Gamma_{22i}(L)^{-1} & -\Gamma_{22i}(L)^{-1}\lambda_{2i}' \\ 0 & 0 & \Psi(L)^{-1} \end{pmatrix}$$

and

$$U_i(L) = \begin{pmatrix} I_r & -((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1} + b_i') & (1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\lambda_{2i}' - \lambda_{1i}' \\ 0 & I_m & 0 \\ 0 & 0 & I_k \end{pmatrix}$$

are invertible.

Substituting (A3) for $C_i(L)$ in the MA representation of $\Delta Z_{i,t}^+$ and pre-multiplying by $U_i(L)$ and $\bar{M}(L)$, where

$$\bar{M}(L) = \begin{pmatrix} I_r & 0 \\ 0 & (1-L)I_{(m+k)} \end{pmatrix},$$

we obtain

$$\bar{M}(L)U_i(L)(1-L)\big(Z_{i,t}^+ - (\pi_i^*)'g_t\big) = (1-L)V_i(L)^{-1}\varepsilon_{i,t}^+.$$

Eliminating $(1-L)$ from both sides and pre-multiplying by $V_i(L)$ yields the following possibly infinite AR representation for $Z_{i,t}^+$

$$V_i(L)\bar{M}(L)U_i(L)\big(Z_{i,t}^+ - (\pi_i^*)'g_t\big) = \varepsilon_{i,t}^+.$$

Using that $\Gamma_{11\cdot 2i}(L)^{-1} = |\Gamma_{11\cdot 2i}(L)|^{-1}adj(\Gamma_{11\cdot 2i}(L))$, $\Gamma_{22i}(L)^{-1} = |\Gamma_{22i}(L)|^{-1}adj(\Gamma_{22i}(L))$, $\Psi(L)^{-1} = |\Psi(L)|^{-1}adj(\Psi(L))$ and $|\Gamma_i^+(L)| = |\Gamma_{11\cdot 2i}(L)||\Gamma_{22i}(L)||\Psi(L)|$, we can recover both the scalar lag polynomial $c_i(L) = |\Gamma_i^+(L)|$ and the lag polynomial matrix $A_i(L)$ given in the theorem. This establishes part (b).

Consider (c). Direct computation of $A_i(1)$ yields

$$A_i(1) = \begin{pmatrix} |\Psi(1)||\Gamma_{22i}(1)|adj(\Gamma_{11\cdot 2i}(1)) \\ -|\Psi(1)|adj(\Gamma_{22i}(1))\Gamma_{21i}(1)adj(\Gamma_{11\cdot 2i}(1)) \\ 0 \end{pmatrix} \begin{pmatrix} I_r & -b_i' & -\lambda_{1i}' \end{pmatrix} = \alpha_i^*\beta_i'.$$

Since

$$C_i(1) = \begin{pmatrix} b_i' & (\lambda_{1i}' + b_i'\lambda_{2i}') \\ I_m & \lambda_{2i}' \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \Gamma_{21i}(1) & \Gamma_{22i}(1) & 0 \\ 0 & 0 & \Psi(1) \end{pmatrix} = \tilde{\beta}_i(\tilde{\alpha}_i^*)',$$

where $\tilde{\alpha}_i^*$ and $\tilde{\beta}_i$ denote the matrices orthogonal to α_i^* and β_i, respectively. It follows that $\beta_i'C_i(1) = 0$ and $C_i(1)\alpha_i = 0$, and so the proof of (c) is complete.

Parts (d) and (e) follow by manipulating of the lag polynomial matrix $A_i(L)$ and rearranging terms, as in Engle and Granger (1987).

A.2 Proof of Theorem 2.

Before we come to the proof of the theorem we need some preliminary results, which are summarized in Lemma 1.

Lemma 1. *Under H_{0i} and Assumptions 1, 2, 4 and 5, as $T \to \infty$*

(a) $T^{-\frac{1}{2}} Z^{+}_{i,t-1} \xrightarrow{w} \Omega_i^{\frac{1}{2}} W_i$,

(b) $T^{-2} \sum_{t=2}^{T} Q_W Z^{+}_{i,t-1} (Q_W Z^{+}_{i,t-1})' \xrightarrow{w} \Omega_i^{\frac{1}{2}} \int W_i W_i' (\Omega_i^{\frac{1}{2}})'$,

(c) $T^{-1} \sum_{t=2}^{T} Q_W Z^{+}_{i,t-1} Q_W \varepsilon_{1\cdot 2i,t} \xrightarrow{w} \sigma_i \Omega_i^{\frac{1}{2}} p_i(W_i)$,

where

$$\begin{aligned} \Omega_i^{\frac{1}{2}} &= \begin{pmatrix} \sigma_i \tilde{B}^{-1}_{11\cdot 2i} & \tilde{B}^{-1}_{11\cdot 2i} M_{1i} M_{2i} \\ \sigma_i \tilde{B}^{-1}_{11\cdot 2i} \rho_i & (M_{2i} + \tilde{B}^{-1}_{11\cdot 2i} \rho_i M_{1i} M_{2i}) \end{pmatrix} \\ &= \begin{pmatrix} \omega_{11i} & \Omega_{12i} \\ \Omega_{21i} & \Omega_{22i} \end{pmatrix}, \end{aligned}$$

where $\tilde{B}_{jji}(L) = I - B_{jji}(L)L$, $B_{jli} = B_{jli}(1)$, $\tilde{B}_{jli} = \tilde{B}_{jli}(1)$, $\tilde{B}^{-1}_{11\cdot 2i} = (\tilde{B}_{11i} - B_{12i}\tilde{B}^{-1}_{22i} B_{21i})^{-1}$, $\rho_i' = \begin{pmatrix} B'_{21i}(\tilde{B}^{-1}_{22i})' & 0 \end{pmatrix}$, $M_{1i} = \begin{pmatrix} B_{12i} & B_{13i} \end{pmatrix}$ *and*

$$M_{2i} = \begin{pmatrix} \tilde{B}^{-1}_{22i} \Sigma^{\frac{1}{2}}_{22i} & \tilde{B}^{-1}_{22i}(B_{23i}\tilde{B}^{-1}_{33i} + \lambda'_{2i}) \\ 0 & \tilde{B}^{-1}_{33i} \end{pmatrix}.$$

Note that $\Omega_{21i} = \omega_{11i}\rho_i$ and define for future use $\omega_{11\cdot 2i} = \omega_{11i} - \omega_{11i}\Omega_{12}\Omega^{-1}_{22i}\rho_i$.

Proof of Lemma 1.

Consider (a). Note that under H_{0i},

$$\begin{pmatrix} \tilde{B}_{11i}(1) & -B_{12i}(1) & -B_{13i}(1) \\ -B_{21i}(1)L & \tilde{B}_{22i}(1) & -B_{23i}(1)L \\ 0 & 0 & \tilde{B}_{33i}(1) \end{pmatrix} \Delta Z^{+}_{i,t} = \begin{pmatrix} \varepsilon_{1\cdot 2i,t} \\ \varepsilon^{*}_{2i,t} \\ \eta_t \end{pmatrix},$$

such that

$$\begin{aligned} Z^{+}_{i,t} &= \begin{pmatrix} \tilde{B}_{11i}(1) & -B_{12i}(1) & -B_{13i}(1) \\ -B_{21i}(1) & \tilde{B}_{22i}(1) & -B_{23i}(1) \\ 0 & 0 & \tilde{B}_{33i}(1) \end{pmatrix}^{-1} \left(\sum_{s=1}^{t} \begin{pmatrix} \varepsilon_{1\cdot 2i,t} \\ \varepsilon^{*}_{2i,t} \\ \eta_t \end{pmatrix} \right. \\ &+ \left. \begin{pmatrix} -\tilde{B}^{+}_{11i}(L) & B^{+}_{12i}(L) & B^{+}_{13i}(L) \\ B^{+}_{21i}(L)L & -\tilde{B}^{+}_{22i}(L) & B^{+}_{23i}(L)L \\ 0 & 0 & -\tilde{B}^{+}_{33i}(L) \end{pmatrix} \Delta Z^{+}_{i,t} \right), \end{aligned} \quad \text{(A4)}$$

where $B^{+}_{jli}(L)$ and $\tilde{B}^{+}_{jli}(L)$ are obtained from the Beveridge-Nelson decompositions of $B_{jli}(L)$ and $\tilde{B}_{jli}(L)$ as $B_{jli}(L) = B_{jli}(1) + B^{+}_{jli}(L)(1-L)$ and $\tilde{B}_{jli}(L) = \tilde{B}_{jli}(1) + \tilde{B}^{+}_{jli}(L)(1-L)$, respectively.

Substituting $\varepsilon^*_{2i,t} = \varepsilon_{2i,t} + \lambda'_{2i}\eta_t$ into (A4) we obtain

$$\begin{aligned} T^{-\frac{1}{2}} Z^+_{i,t} &= \begin{pmatrix} \tilde{B}_{11i}(1) & -B_{12i}(1) & -B_{13i}(1) \\ -B_{21i}(1) & \tilde{B}_{22i}(1) & -B_{23i}(1) \\ 0 & 0 & \tilde{B}_{33i}(1) \end{pmatrix}^{-1} \begin{pmatrix} \sigma_i & 0 & 0 \\ 0 & \Sigma^{\frac{1}{2}}_{22i} & \lambda'_{2i} \\ 0 & 0 & I_k \end{pmatrix} \\ &\times \quad T^{-\frac{1}{2}} \sum_{s=1}^{t} \begin{pmatrix} \sigma_i^{-1}\varepsilon_{1\cdot 2i,t} \\ \Sigma^{-\frac{1}{2}}_{22i}\varepsilon_{2i,t} \\ \eta_t \end{pmatrix} + o_p(1), \end{aligned}$$

such that

$$T^{-\frac{1}{2}} Z^+_{i,t} \xrightarrow{w} \Omega_i^{\frac{1}{2}} W_i \tag{A5}$$

as $T \to \infty$, proving (a).

Now, by using the rules for projections, $\sum_{t=2}^{T} Q_W Z^+_{i,t-1}(Q_W Z^+_{i,t-1})'$ can be written as

$$\begin{aligned} \sum_{t=2}^{T} Q_W Z^+_{i,t-1}(Q_W Z^+_{i,t-1})' &= \sum_{t=2}^{T} Z^+_{i,t-1}(Z^+_{i,t-1})' \\ &- \sum_{t=2}^{T} Z^+_{i,t-1}W'_{i,t}\left(\sum_{t=2}^{T} W_{i,t}W'_{i,t}\right)^{-1}\sum_{t=2}^{T} W_{i,t}(Z^+_{i,t-1})'. \end{aligned} \tag{A6}$$

By Lemma 2.1 of Park and Phillips (1989), $\sum_{t=2}^{T} Z^+_{i,t-1}W'_{i,t} = O_p(T)$, $\sum_{t=2}^{T} Z^+_{i,t-1}(Z^+_{i,t-1})' = O_p(T^2)$ and $\sum_{t=2}^{T} W_{i,t}W'_{i,t} = O_p(T)$ such that (A6) reduces to

$$\begin{aligned} T^{-2}\sum_{t=2}^{T} Q_W Z^+_{i,t-1}(Q_W Z^+_{i,t-1})' &= T^{-2}\sum_{t=2}^{T} Z^+_{i,t-1}(Z^+_{i,t-1})' + T^{-2}O_p(T)O_p(T^{-1})O_p(T) \\ &= T^{-2}\sum_{t=2}^{T} Z^+_{i,t-1}(Z^+_{i,t-1})' + O_p(T^{-1}), \end{aligned}$$

where we can make use of (a) to show that as $T \to \infty$

$$T^{-2}\sum_{t=2}^{T} Z^+_{i,t-1}(Z^+_{i,t-1})' \xrightarrow{w} \Omega_i^{\frac{1}{2}} \int W_i W'_i (\Omega_i^{\frac{1}{2}})'.$$

This proves (b).

Finally, consider (c). By definition,

$$\sum_{t=2}^{T} Q_W Z^+_{i,t-1}(Q_W \varepsilon_{1\cdot 2i,t}) = \sum_{t=2}^{T} Z^+_{i,t-1}\varepsilon_{1\cdot 2i,t} - \sum_{t=2}^{T} Z^+_{i,t-1}W'_{i,t}\left(\sum_{t=2}^{T} W_{i,t}W'_{i,t}\right)^{-1}\sum_{t=2}^{T} W_{i,t}\varepsilon_{1\cdot 2i,t},$$

where $\sum_{t=2}^{T} W_{i,t}\varepsilon_{1\cdot 2i,t} = O_p(\sqrt{T})$. Thus, by using the same arguments as above,

$$T^{-1}\sum_{t=2}^{T} Q_W Z^+_{i,t-1}(Q_W \varepsilon_{1\cdot 2i,t}) = T^{-1}\sum_{t=2}^{T} Z^+_{i,t-1}\varepsilon_{1\cdot 2i,t} + O_p(T^{-\frac{1}{2}}), \tag{A7}$$

where the limit of the first term on the right-hand side is given by

$$T^{-1}\sum_{t=2}^{T} Z^{+}_{i,t-1}\varepsilon_{1\cdot 2i,t} \xrightarrow{w} \sigma_i\Omega_i^{\frac{1}{2}}\int W_i dW_{1i} = \sigma_i\Omega_i^{\frac{1}{2}}p_i(W_i).$$

This establishes (c), and hence the proof of Lemma 1 is complete. □

Now, since under the null hypothesis,

$$\Delta(Q_W Y_{i,t}) = \delta'_{1i}(Q_W Z^{+}_{i,t-1}) + Q_W\varepsilon_{1\cdot 2i,t} = Q_W\varepsilon_{1\cdot 2i,t},$$

we have

$$\hat{\delta}_{1i} = \left(\sum_{t=2}^{T} Q_W Z^{+}_{i,t-1}(Q_W Z^{+}_{i,t-1})'\right)^{-1}\sum_{t=2}^{T} Q_W Z^{+}_{i,t-1}(Q_W\varepsilon_{1\cdot 2i,t}).$$

From Lemma 1 (b) and (c) we have that

$$T\hat{\delta}_{1i} \xrightarrow{w} \sigma_i\left(\Omega_i^{\frac{1}{2}}\int W_i W_i'(\Omega_i^{\frac{1}{2}})'\right)^{-1}\Omega_i^{\frac{1}{2}}\int W_i dW_{1i} = \sigma_i(\Omega_i^{-\frac{1}{2}})'P_i(W_i). \tag{A8}$$

Similarly, under the null the Wald statistic is given by

$$w_{\hat{\delta}_{1i}} = \hat{\sigma}_i^{-2}\sum_{t=2}^{T} Q_W\varepsilon_{1\cdot 2i,t}(Q_W Z^{+}_{i,t})'\left(\sum_{t=2}^{T} Q_W Z^{+}_{i,t-1}(Q_W Z^{+}_{i,t-1})'\right)^{-1}\sum_{t=2}^{T} Q_W Z^{+}_{i,t-1}(Q_W\varepsilon_{1\cdot 2i,t}).$$

Consider $\hat{\sigma}_i^2 = T^{-1}\sum_{t=2}^{T}(\Delta(Q_W Y_{i,t}) - \hat{\delta}'_{1i}(Q_W Z^{+}_{i,t-1}))^2$. By making use of Lemma 1, and the fact that under the null, $\Delta(Q_W Y_{i,t}) = Q_W\varepsilon_{1\cdot 2i,t}$, we get

$$\begin{aligned}
\hat{\sigma}_i^2 &= T^{-1}\sum_{t=2}^{T}(\Delta(Q_W Y_{i,t} - \hat{\delta}_{1i}(Q_W Z^{+}_{i,t-1}))^2 \\
&= T^{-1}\sum_{t=2}^{T}(Q_W\varepsilon_{1\cdot 2i,t})^2 - 2\hat{\delta}'_{1i}T^{-1}\sum_{t=2}^{T} Q_W Z^{+}_{i,t-1}(Q_W\varepsilon_{1\cdot 2i,t}) \\
&+ \hat{\delta}'_{1i}T^{-1}\sum_{t=2}^{T} Q_W Z^{+}_{i,t-1}(Q_W Z^{+}_{i,t-1})'\hat{\delta}_{1i} \\
&= T^{-1}\sum_{t=2}^{T}(Q_W\varepsilon_{1\cdot 2i,t})^2 + O_p(T^{-1})O_p(1) + T^{-1}O_p(T^{-1})O_p(T^2)O_p(T^{-1}) \\
&= T^{-1}\sum_{t=2}^{T}(Q_W\varepsilon_{1\cdot 2i,t})^2 + O_p(T^{-1}).
\end{aligned} \tag{A9}$$

As for the first term on the right-hand side, we have

$$\begin{aligned}
\sum_{t=2}^{T}(Q_W\varepsilon_{1\cdot 2i,t})^2 &= \sum_{t=2}^{T}\varepsilon^2_{1\cdot 2i,t} - \sum_{t=2}^{T}\varepsilon_{1\cdot 2i,t}W'_{i,t}\left(\sum_{t=2}^{T} W_{i,t}W'_{i,t}\right)^{-1}\sum_{t=2}^{T} W_{i,t}\varepsilon_{1\cdot 2i,t} \\
&= \sum_{t=2}^{T}\varepsilon^2_{1\cdot 2i,t} + O_p(\sqrt{T})O_p(T^{-1})O_p(\sqrt{T}).
\end{aligned}$$

Hence, by inserting this into (A9), and then taking the limit as $T \to \infty$, we obtain

$$\hat{\sigma}_i^2 = T^{-1} \sum_{t=2}^{T} \varepsilon_{1\cdot 2i,t}^2 + O_p(T^{-1}) \xrightarrow{p} \sigma_i^2. \tag{A10}$$

Combining this result with Lemma 1 (b) and (c) we get the following limit as $T \to \infty$

$$\begin{aligned} w_{\hat{\delta}_{1i}} &\xrightarrow{w} \sigma_i^{-2}\sigma_i \int dW_{1i} W_i' (\Omega_i^{\frac{1}{2}})' \left(\Omega_i^{\frac{1}{2}} \int W_i W_i' (\Omega_i^{\frac{1}{2}})' \right)^{-1} \sigma_i \Omega_i^{\frac{1}{2}} \int W_i dW_{1i} \\ &= p_i(W_i)' P_i(W_i), \end{aligned} \tag{A11}$$

which establishes part (a) of the theorem.

Consider (b). Under the null hypothesis, $\Delta(Q_S Y_{i,t}) = Q_S \varepsilon_{1\cdot 2i,t}$. By using this result, (A8) and the rules for partitioned regressions, we obtain as $T \to \infty$

$$\begin{aligned} T\hat{\alpha}_{1i} &= \left(T^{-2} \sum_{t=2}^{T} (Q_S Y_{i,t-1})^2 \right)^{-1} T^{-1} \sum_{t=2}^{T} Q_S Y_{i,t-1} (Q_S \varepsilon_{1\cdot 2i,t}) \\ &\xrightarrow{w} \sigma_i \omega_{11\cdot 2i}^{-1} P_i(U_i) + \omega_{11\cdot 2i}^{-1} \omega_{11i} \rho_i' (\Omega_{22i}')^{-1} P_i(W_{2\cdot i})(P_i(U_i) - 1) = d_i. \end{aligned} \tag{A12}$$

Next, consider

$$\mathrm{var}(\hat{\alpha}_{1i}) = \hat{\sigma}_i^2 \left(T^{-2} \sum_{t=2}^{T} (Q_S Y_{i,t-1})^2 \right)^{-1}.$$

We have already shown that $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ as $T \to \infty$. From this result and arguments similar to those used in the proof of Lemma 1 we obtain as $T \to \infty$

$$\begin{aligned} T^2 \mathrm{var}(\hat{\alpha}_{1i}) &\xrightarrow{w} \sigma_i^2 \omega_{11\cdot 2i}^{-2} V(U_i) \\ &+ \sigma_i^2 \omega_{11\cdot 2i}^{-2} \omega_{11i} V(U_i) \left(\rho_i' (\Omega_{22i}')^{-1} P_i(W_{2\cdot i}) + V(W_{2\cdot i}) p_i(W_{2\cdot i})' \Omega_{22i}^{-1} \rho_i \right) \\ &+ \sigma_i^2 \omega_{11\cdot 2i}^{-2} \omega_{11i}^2 \rho_i' (\Omega_{22i}')^{-1} \\ &\times \left(V(W_{2\cdot i}) + P_i(W_{2\cdot i}) V(U_i) V(W_{2\cdot i}) p_i(W_{2\cdot i})' \right) \Omega_{22i}^{-1} \rho_i = D_i. \end{aligned} \tag{A13}$$

The proof is completed by noting that

$$D_{i,\tau}^{\Omega} = \lim_{T \to \infty} \frac{T\hat{\alpha}_{1i}}{\sqrt{T^2 \mathrm{var}(\hat{\alpha}_{1i})}} = \frac{d_i}{\sqrt{D_i}}. \tag{A14}$$

□

A.2.1 Proof of Corollary 1

If $X_{i,t}$ is strongly exogenous, $B_{21i} = 0$ such that $\rho_i = 0$. Thus, (A14) simplifies to

$$D_{i,\tau} = \frac{d_i}{\sqrt{D_i}} = \frac{\tilde{B}_{11i} P_i(U_i)}{\sqrt{\tilde{B}_{11i}^2 V(U_i)}} = \frac{P_i(U_i)}{\sqrt{V(U_i)}}. \tag{A15}$$

This completes the proof. □

A.3 Proof of Theorem 3

We begin with the following lemma.

Lemma 2. *Under H_{0i} and Assumptions 1 to 6, as $N, T \to \infty$*

(a) $T^{-\frac{1}{2}}\overline{Z}_t \xrightarrow{w} M_3 W_3$,

(b) $T^{-\frac{1}{2}}\tilde{Z}^{+}_{i,t-1} \xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} W_i$,

(c) $T^{-2}\sum_{t=2}^{T} Q_{\tilde{W}}\tilde{Z}^{+}_{i,t}(Q_{\tilde{W}}\tilde{Z}^{+}_{i,t})' \xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}}\left(\int W_i W_i'\right)(\tilde{\Omega}_i^{\frac{1}{2}})'$,

(d) $T^{-1}\sum_{t=2}^{T} Q_{\tilde{W}}\tilde{Z}^{+}_{i,t-1}(Q_{\tilde{W}}\tilde{\varepsilon}_{1\cdot 2i,t}) \xrightarrow{w} \sigma_i\tilde{\Omega}_i^{\frac{1}{2}} p_i(W_i)$,

where $M_3 = \lim_{N\to\infty}\overline{M}_3$, $\overline{M}_3 = \frac{1}{N}\sum_{i=1}^{N} M_{3i}$,

$$M_{3i} = \begin{pmatrix} \tilde{B}^{-1}_{11\cdot 2i}\big(B_{12i}\tilde{B}^{-1}_{22i}(B_{23i}\tilde{B}^{-1}_{33i} + \lambda'_{2i}) + B_{13i}\tilde{B}^{-1}_{33i}\big) \\ \tilde{B}^{-1}_{22i}(B_{23i}\tilde{B}^{-1}_{33i} + \lambda'_{2i}) + \tilde{B}^{-1}_{22i}B_{21i}\tilde{B}^{-1}_{11\cdot 2i}\big(B_{12i}\tilde{B}^{-1}_{22i}(B_{23i}\tilde{B}^{-1}_{33i} + \lambda'_{2i}) + B_{13i}\tilde{B}^{-1}_{33i}\big) \end{pmatrix},$$

$$M_{4i} = \begin{pmatrix} \tilde{B}^{-1}_{22i}\Sigma^{\frac{1}{2}}_{22i} & \tilde{B}^{-1}_{22i}(B_{23i}\tilde{B}^{-1}_{33i} + \lambda'_{2i}) \\ 0 & M_3 \end{pmatrix},$$

$\tilde{\rho}_i = \begin{pmatrix} B'_{21i}(\tilde{B}^{-1}_{22i})' & 0 \end{pmatrix}$ *and*

$$\begin{aligned} \tilde{\Omega}_i^{\frac{1}{2}} &= \begin{pmatrix} \sigma_i\tilde{B}^{-1}_{11\cdot 2i} & \tilde{B}^{-1}_{11\cdot 2i}M_{1i}M_{2i} \\ \sigma_i\tilde{B}^{-1}_{11\cdot 2i}\tilde{\phi}_i & (M_{4i} + \tilde{B}^{-1}_{11\cdot 2i}\tilde{\phi}_i M_{1i}M_{2i}) \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\omega}_{11i} & \tilde{\Omega}_{12i} \\ \tilde{\Omega}_{21i} & \tilde{\Omega}_{22i} \end{pmatrix}. \end{aligned}$$

Furthermore, we have $\tilde{\Omega}_{21i} = \tilde{\omega}_{11i}\tilde{\rho}_i$ and we define $\tilde{\omega}_{11\cdot 2i} = \tilde{\omega}_{11i} - \tilde{\omega}_{11i}\tilde{\Omega}_{12i}\tilde{\Omega}^{-1}_{22i}\tilde{\rho}_i$.

Proof of Lemma 2

Letting $\varphi_i = \tilde{B}^{-1}_{11\cdot 2i}B_{12i}(1)\tilde{B}_{22i}(1)^{-1}$, we have

$$\begin{aligned} T^{-\frac{1}{2}}\overline{Z}_t &= \overline{M}_3 T^{-\frac{1}{2}}\sum_{s=1}^{t}\varepsilon_{3s} + \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \tilde{B}^{-1}_{11\cdot 2i} & \varphi_i \\ \varphi'_i & (\tilde{B}_{22i}(1)^{-1} + \tilde{B}_{22i}(1)^{-1}B_{21i}(1)\varphi_i) \end{pmatrix} \\ &\times \; T^{-\frac{1}{2}}\sum_{s=1}^{t}\begin{pmatrix} \varepsilon_{1\cdot 2i,s} \\ \varepsilon_{2i,s} \end{pmatrix} + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

from which it follows that as $N, T \to \infty$

$$\begin{aligned} T^{-\frac{1}{2}}\overline{Z}_t &= \overline{M}_3 T^{-\frac{1}{2}}\sum_{s=1}^{t}\varepsilon_{3s} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\xrightarrow{w} M_3 W_3. \end{aligned} \tag{A16}$$

This proves (a).

Moreover, by combining (a) with (A5), we find that as N, $T \to \infty$

$$T^{-\frac{1}{2}} \tilde{Z}_{i,t}^{+} \xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} W_i,$$

which proves (b).

Analogous to the prove of Lemma 1, we have

$$\begin{aligned}
\sum_{t=2}^{T} Q_{\tilde{W}} \tilde{Z}_{i,t-1}^{+} (Q_{\tilde{W}} \tilde{Z}_{i,t-1}^{+})' &= \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} (\tilde{Z}_{i,t-1}^{+})' - \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} (\tilde{W}_{i,t})' \left(\sum_{t=2}^{T} \tilde{W}_{i,t} (\tilde{W}_{i,t})' \right)^{-1} \\
&\times \sum_{t=2}^{T} \tilde{W}_{i,t} (\tilde{Z}_{i,t-1}^{+})' \\
&= \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} (\tilde{Z}_{i,t-1}^{+})' + O_p(T).
\end{aligned}$$

Combining this with (b) we obtain as N, $T \to \infty$

$$\begin{aligned}
T^{-2} \sum_{t=2}^{T} Q_{\tilde{W}} \tilde{Z}_{i,t-1}^{+} (Q_{\tilde{W}} \tilde{Z}_{i,t-1}^{+})' &= T^{-2} \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} (\tilde{Z}_{i,t-1}^{+})' + O_p(T^{-1}) \\
&\xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} \left(\int W_i W_i' \right) (\tilde{\Omega}_i^{\frac{1}{2}})'. \qquad \text{(A17)}
\end{aligned}$$

This proves (c).

Finally,

$$\begin{aligned}
\sum_{t=2}^{T} Q_{\tilde{W}} \tilde{Z}_{i,t-1}^{+} (Q_{\tilde{W}} \tilde{\varepsilon}_{1\cdot 2i,t}) &= \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} \tilde{\varepsilon}_{1\cdot 2i,t} - \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} (\tilde{W}_{i,t})' \left(\sum_{t=2}^{T} \tilde{W}_{i,t} (\tilde{W}_{i,t})' \right)^{-1} \\
&\times \sum_{t=2}^{T} \tilde{W}_{i,t} (\tilde{\varepsilon}_{1\cdot 2i,t}) \\
&= \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} \tilde{\varepsilon}_{1\cdot 2i,t} + O_p(T) O_p(T^{-1}) O_p(\sqrt{T}).
\end{aligned}$$

Thus,

$$\begin{aligned}
T^{-1} \sum_{t=2}^{T} Q_{\tilde{W}} \tilde{Z}_{i,t-1}^{+} (Q_{\tilde{W}} \tilde{\varepsilon}_{1\cdot 2i,t}) &= T^{-1} \sum_{t=2}^{T} \tilde{Z}_{i,t-1}^{+} \tilde{\varepsilon}_{1\cdot 2i,t} + O_p(T^{-\frac{1}{2}}) \\
&\xrightarrow{w} \sigma_i \tilde{\Omega}_i^{\frac{1}{2}} p_i(W_i) \qquad \text{(A18)}
\end{aligned}$$

as N, $T \to \infty$. This proves (d) and hence the proof of Lemma 2 is complete. □

The proof of Theorem 2 follows similar arguments as the proof of Theorem 3. However, if $k < m + 1$, $\tilde{\Omega}_i^{\frac{1}{2}}$ and $\tilde{\Omega}_{22i}$ are no longer square matrices such that we have to make use of generalized inverse in that case.

The Wald statistic $w_{\tilde{\delta}_{1i}}$ is given by

$$w_{\tilde{\delta}_{1i}} = \breve{\sigma}_i^{-2} \sum_{t=2}^{T} Q_{\tilde{W}} \tilde{\varepsilon}_{1\cdot 2i,t} (Q_{\tilde{W}} \tilde{Z}^+_{i,t})' \left(\sum_{t=2}^{T} Q_{\tilde{W}} \tilde{Z}^+_{i,t-1} (Q_{\tilde{W}} \tilde{Z}^+_{i,t-1})' \right)^{-1} \tag{A19}$$

$$\times \sum_{t=2}^{T} Q_{\tilde{W}} \tilde{Z}^+_{i,t-1} (Q_{\tilde{W}} \tilde{\varepsilon}_{1\cdot 2i,t}), \tag{A20}$$

where $\tilde{\sigma}_i^2$ is $\hat{\sigma}_i^2$ with $Q_{\tilde{W}}$ in place of Q_W. By using the same steps as for $\hat{\sigma}_i^2$ in Theorem 2, we obtain $\tilde{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ as $N, T \to \infty$. This result, together with Lemma 2 (c) and (d), implies that as $N, T \to \infty$

$$\begin{aligned} w_{\tilde{\delta}_{1i}} &\xrightarrow{w} \sigma_i^{-2} \sigma_i \int dW_{1i} W_i' (\tilde{\Omega}_i^{\frac{1}{2}})' \left(\tilde{\Omega}_i^{\frac{1}{2}} \int W_i W_i' (\tilde{\Omega}_i^{\frac{1}{2}})' \right)^{-1} \sigma_i \tilde{\Omega}_i^{\frac{1}{2}} \int W_i dW_{1i} \\ &= p_i(W_i)' P_i(W_i), \end{aligned} \tag{A21}$$

which establishes the required result for (a).

Furthermore, similarly to the prove of Theorem 1, by the rules for partitioned regressions, $T\tilde{\alpha}_{1i} \xrightarrow{w} \tilde{d}_i$ and $T^2 \mathrm{var}(\tilde{\alpha}_{1i}) \xrightarrow{w} \tilde{D}_i$ as N, $T \to \infty$, where $\tilde{d}_i$ and $\tilde{D}_i$ are defined similarly to d_i and D_i above, but replacing ω_{11i}, $\omega_{11\cdot 2i}$ and Ω_{22i} with $\tilde{\omega}_{11i}$, $\tilde{\omega}_{11\cdot 2i}$ and $\tilde{\Omega}_{22i}$ respectively. This yields the required result for (b). □

A.3.1 Proof of Corollary 1

The proof of Corollary 2 is completed by noting that $\tilde{\rho}_i = 0$ if $X_{i,t}$ is strongly exogenous. □

A.4 Tables

Table 1: Critical values and moments for the individual t-tests.

		Critical values			Moments	
Model	m	10%	5%	1%	$\mathsf{E}(D_\tau)$	$\text{var}(D_\tau)$
1	1	−2.985	−3.315	−3.932	−1.709	1.069
	2	−3.484	−3.819	−4.434	−2.212	1.044
	3	−3.883	−4.219	−4.848	−2.617	1.026
	4	−4.233	−4.570	−5.191	−2.965	1.020
	5	−4.538	−4.876	−5.503	−3.272	1.012
2	1	−3.426	−3.744	−4.339	−2.250	0.884
	2	−3.845	−4.168	−4.775	−2.644	0.915
	3	−4.199	−4.528	−5.138	−2.985	0.931
	4	−4.512	−4.841	−5.454	−3.287	0.943
	5	−4.792	−5.123	−5.747	−3.564	0.947
3	1	−3.814	−4.122	−4.697	−2.704	0.779
	2	−4.175	−4.488	−5.078	−3.024	0.837
	3	−4.494	−4.815	−5.411	−3.316	0.872
	4	−4.780	−5.103	−5.703	−3.589	0.892
	5	−5.043	−5.370	−5.973	−3.841	0.904

Notes: Model 1 refers to the specification with no deterministic component, while Models 2 and 3 refer to the specifications with an unrestricted constant, and unrestricted constant and trend, respectively. The value m refers to the number of regressors contained in $X_{i,t}$.

Table 2: Critical values and moments for the individual Wald tests.

		Critical values			Moments	
Model	m	10%	5%	1%	$\mathsf{E}(D_w)$	$\mathrm{var}(D_w)$
1	1	12.209	14.291	18.726	6.979	15.188
	2	17.399	19.839	24.913	10.937	23.438
	3	22.344	25.010	30.634	14.872	31.381
	4	27.108	30.061	36.132	18.785	39.317
	5	31.795	34.966	41.435	22.709	47.043
2	1	14.821	17.081	21.870	8.944	19.467
	2	19.870	22.460	27.817	12.886	27.601
	3	24.750	27.571	33.400	16.833	35.554
	4	29.484	32.542	38.789	20.756	43.392
	5	34.076	37.329	43.941	24.639	50.867
3	1	17.525	19.940	24.973	11.091	23.396
	2	22.424	25.113	30.674	14.988	31.266
	3	27.190	30.127	36.200	18.891	39.200
	4	31.840	34.992	41.404	22.767	46.941
	5	36.389	39.768	46.581	26.639	54.563
4	1	15.769	18.012	22.789	9.964	18.782
	2	20.781	23.337	28.680	13.898	26.598
	3	25.629	28.422	34.305	17.824	34.284
	4	30.368	33.430	39.625	21.756	42.014
	5	34.995	38.236	44.840	25.648	49.671
5	1	18.412	20.800	25.830	12.093	22.321
	2	23.297	25.968	31.550	15.982	30.128
	3	28.084	31.016	37.034	19.888	37.981
	4	32.708	35.839	42.256	23.757	45.521
	5	37.293	40.612	47.334	27.628	53.097

Notes: Models 4 and 5 refer to the specifications with a constant, and constant and trend in the error correction term, respectively. See Table 1 for an explanation of the remaining features of the table.

Table 3: Critical values for the panel t-tests.

	$m=1$			$m=2$			$m=3$			$m=4$			$m=5$		
N	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Model 1															
10	−2.181	−2.300	−2.519	−2.655	−2.771	−2.976	−3.060	−3.192	−3.394	−3.400	−3.519	−3.745	−3.688	−3.812	−4.013
15	−2.131	−2.232	−2.417	−2.590	−2.687	−2.866	−2.988	−3.085	−3.275	−3.322	−3.418	−3.597	−3.628	−3.725	−3.904
20	−2.103	−2.200	−2.380	−2.552	−2.644	−2.820	−2.941	−3.024	−3.182	−3.281	−3.368	−3.534	−3.583	−3.677	−3.832
25	−2.067	−2.163	−2.318	−2.531	−2.617	−2.779	−2.924	−3.006	−3.148	−3.254	−3.332	−3.472	−3.557	−3.636	−3.789
30	−2.048	−2.133	−2.287	−2.513	−2.588	−2.733	−2.899	−2.979	−3.114	−3.239	−3.305	−3.448	−3.536	−3.607	−3.744
40	−2.035	−2.121	−2.267	−2.493	−2.566	−2.706	−2.871	−2.947	−3.078	−3.208	−3.271	−3.398	−3.509	−3.574	−3.701
50	−2.019	−2.094	−2.234	−2.476	−2.545	−2.669	−2.859	−2.921	−3.048	−3.187	−3.254	−3.371	−3.491	−3.556	−3.672
70	−2.012	−2.087	−2.212	−2.457	−2.520	−2.627	−2.838	−2.897	−3.004	−3.169	−3.222	−3.332	−3.470	−3.525	−3.623
100	−1.995	−2.065	−2.190	−2.440	−2.505	−2.616	−2.822	−2.876	−2.975	−3.152	−3.204	−3.297	−3.451	−3.503	−3.594
Model 2															
10	−2.658	−2.772	−2.971	−3.056	−3.167	−3.377	−3.393	−3.513	−3.728	−3.692	−3.804	−4.010	−3.966	−4.082	−4.284
15	−2.603	−2.698	−2.866	−2.993	−3.083	−3.265	−3.325	−3.420	−3.599	−3.624	−3.714	−3.887	−3.897	−3.997	−4.179
20	−2.568	−2.653	−2.796	−2.948	−3.033	−3.196	−3.283	−3.370	−3.525	−3.588	−3.670	−3.820	−3.862	−3.941	−4.096
25	−2.544	−2.623	−2.773	−2.925	−3.000	−3.139	−3.258	−3.331	−3.468	−3.560	−3.643	−3.784	−3.832	−3.906	−4.066
30	−2.530	−2.601	−2.735	−2.909	−2.981	−3.120	−3.238	−3.305	−3.438	−3.538	−3.607	−3.746	−3.810	−3.882	−4.008
40	−2.504	−2.574	−2.694	−2.885	−2.949	−3.071	−3.221	−3.285	−3.404	−3.515	−3.579	−3.702	−3.785	−3.848	−3.956
50	−2.492	−2.554	−2.672	−2.863	−2.924	−3.032	−3.200	−3.259	−3.374	−3.496	−3.551	−3.658	−3.774	−3.830	−3.925
70	−2.477	−2.536	−2.650	−2.848	−2.903	−3.016	−3.181	−3.234	−3.332	−3.475	−3.523	−3.620	−3.751	−3.799	−3.894
100	−2.458	−2.517	−2.611	−2.835	−2.886	−2.983	−3.162	−3.213	−3.301	−3.459	−3.507	−3.584	−3.729	−3.776	−3.862
Model 3															
10	−3.085	−3.188	−3.382	−3.410	−3.514	−3.707	−3.709	−3.820	−4.021	−3.983	−4.098	−4.299	−4.231	−4.347	−4.555
15	−3.026	−3.112	−3.276	−3.351	−3.441	−3.616	−3.644	−3.737	−3.904	−3.921	−4.009	−4.185	−4.173	−4.263	−4.428
20	−2.980	−3.060	−3.212	−3.306	−3.389	−3.531	−3.595	−3.677	−3.827	−3.878	−3.955	−4.108	−4.130	−4.208	−4.344
25	−2.965	−3.041	−3.189	−3.286	−3.361	−3.490	−3.577	−3.657	−3.784	−3.853	−3.927	−4.064	−4.101	−4.172	−4.299
30	−2.948	−3.013	−3.130	−3.269	−3.337	−3.460	−3.563	−3.632	−3.756	−3.831	−3.898	−4.025	−4.083	−4.147	−4.284
40	−2.921	−2.982	−3.095	−3.244	−3.306	−3.420	−3.535	−3.594	−3.706	−3.805	−3.865	−3.972	−4.053	−4.113	−4.223
50	−2.906	−2.964	−3.067	−3.227	−3.282	−3.395	−3.519	−3.576	−3.681	−3.785	−3.845	−3.953	−4.037	−4.091	−4.202
70	−2.892	−2.946	−3.045	−3.210	−3.258	−3.355	−3.500	−3.548	−3.645	−3.764	−3.817	−3.906	−4.015	−4.063	−4.148
100	−2.875	−2.925	−3.010	−3.193	−3.237	−3.319	−3.485	−3.526	−3.606	−3.749	−3.793	−3.881	−3.996	−4.040	−4.119

Notes: See Table 1 for an explanation of the various features of this table.

Table 4: Critical values for the pooled Wald tests.

	$m = 1$			$m = 2$			$m = 3$			$m = 4$			$m = 5$		
N	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
							Model 1								
10	8.724	9.296	10.478	13.053	13.689	14.974	17.246	17.988	19.383	21.394	22.227	23.772	25.514	26.366	28.032
15	8.466	8.927	9.785	12.710	13.244	14.282	16.895	17.511	18.625	20.965	21.708	23.022	25.138	25.843	27.185
20	8.311	8.709	9.419	12.521	13.016	13.882	16.671	17.234	18.186	20.728	21.327	22.422	24.829	25.430	26.690
25	8.209	8.570	9.275	12.403	12.858	13.681	16.525	16.979	17.872	20.591	21.114	22.067	24.627	25.218	26.224
30	8.164	8.505	9.211	12.304	12.698	13.571	16.406	16.855	17.690	20.473	21.013	21.908	24.488	25.030	26.023
40	8.087	8.389	8.947	12.163	12.523	13.149	16.219	16.649	17.425	20.293	20.743	21.635	24.265	24.731	25.578
50	7.989	8.307	8.877	12.097	12.430	13.068	16.152	16.528	17.258	20.187	20.627	21.347	24.168	24.595	25.431
70	7.921	8.207	8.719	11.993	12.310	12.913	16.036	16.363	16.975	20.032	20.391	21.065	24.032	24.409	25.102
100	7.883	8.135	8.593	11.911	12.197	12.696	15.913	16.219	16.788	19.867	20.205	20.787	23.891	24.227	24.790
							Model 2								
10	10.881	11.499	12.781	15.158	15.862	17.156	19.386	20.153	21.721	23.543	24.373	26.042	27.693	28.571	30.120
15	10.564	11.051	12.045	14.801	15.344	16.447	18.980	19.603	20.786	23.075	23.722	25.053	27.128	27.853	29.419
20	10.438	10.875	11.728	14.587	15.137	16.077	18.695	19.266	20.364	22.768	23.332	24.393	26.823	27.540	28.805
25	10.275	10.667	11.487	14.394	14.871	15.684	18.506	19.004	20.033	22.576	23.117	24.145	26.644	27.231	28.323
30	10.212	10.584	11.262	14.348	14.767	15.499	18.433	18.897	19.819	22.432	22.910	23.860	26.502	27.063	28.126
40	10.103	10.437	11.036	14.194	14.530	15.248	18.247	18.648	19.446	22.307	22.741	23.542	26.240	26.698	27.550
50	10.042	10.364	10.940	14.083	14.423	15.042	18.133	18.526	19.200	22.156	22.564	23.359	26.136	26.580	27.380
70	9.897	10.183	10.736	14.008	14.322	14.901	17.980	18.332	19.010	22.010	22.363	22.998	25.968	26.355	27.020
100	9.845	10.096	10.585	13.866	14.150	14.727	17.910	18.206	18.755	21.898	22.213	22.813	25.794	26.142	26.754
							Model 3								
10	13.240	13.882	15.123	17.414	18.194	19.572	21.541	22.401	24.024	25.675	26.537	28.143	29.711	30.684	32.492
15	12.837	13.403	14.406	17.002	17.607	18.702	21.051	21.678	22.975	25.144	25.879	27.175	29.221	30.010	31.559
20	12.608	13.072	13.969	16.722	17.273	18.381	20.798	21.358	22.425	24.943	25.624	26.776	28.863	29.572	30.719
25	12.461	12.903	13.748	16.568	17.023	17.870	20.658	21.168	22.180	24.700	25.240	26.273	28.699	29.277	30.319
30	12.398	12.793	13.514	16.500	16.929	17.811	20.499	20.998	21.938	24.469	25.007	25.956	28.509	29.090	30.142
40	12.254	12.585	13.239	16.316	16.674	17.393	20.372	20.804	21.575	24.347	24.806	25.699	28.367	28.875	29.820
50	12.188	12.514	13.124	16.229	16.585	17.313	20.231	20.617	21.389	24.190	24.626	25.453	28.173	28.620	29.488
70	12.099	12.389	12.950	16.096	16.429	17.017	20.103	20.456	21.109	24.035	24.393	25.102	27.987	28.421	29.132
100	11.977	12.243	12.721	15.993	16.268	16.778	19.960	20.265	20.876	23.908	24.229	24.895	27.837	28.203	28.849

Continued overleaf

Table 4: Continued.

	$m=1$			$m=2$			$m=3$			$m=4$			$m=5$		
N	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
							Model 4								
10	11.913	12.551	13.708	16.139	16.817	18.132	20.376	21.180	22.669	24.463	25.285	26.872	28.600	29.452	31.153
15	11.541	12.033	12.903	15.822	16.386	17.485	19.900	20.509	21.757	24.061	24.685	26.021	28.135	28.969	30.308
20	11.386	11.789	12.690	15.554	16.047	17.081	19.640	20.200	21.258	23.734	24.360	25.445	27.774	28.407	29.622
25	11.294	11.689	12.471	15.404	15.858	16.751	19.497	19.989	20.886	23.544	24.079	25.155	27.629	28.202	29.327
30	11.215	11.556	12.241	15.302	15.715	16.543	19.373	19.840	20.720	23.449	23.932	24.891	27.490	28.024	28.947
40	11.082	11.403	12.080	15.218	15.591	16.334	19.227	19.605	20.445	23.257	23.714	24.520	27.278	27.748	28.660
50	11.014	11.347	11.861	15.098	15.465	16.150	19.124	19.499	20.163	23.163	23.566	24.348	27.135	27.565	28.413
70	10.944	11.223	11.755	14.955	15.264	15.857	18.967	19.299	19.932	22.979	23.323	24.069	26.975	27.361	28.086
100	10.880	11.137	11.633	14.886	15.175	15.696	18.874	19.179	19.719	22.848	23.208	23.781	26.810	27.139	27.771
							Model 5								
10	14.166	14.818	16.222	18.334	19.047	20.452	22.525	23.277	24.967	26.616	27.443	29.169	30.711	31.665	33.368
15	13.813	14.310	15.340	17.970	18.555	19.613	22.095	22.788	24.037	26.075	26.827	28.317	30.177	30.902	32.330
20	13.600	14.088	14.941	17.695	18.183	19.263	21.788	22.358	23.424	25.801	26.437	27.596	29.846	30.514	31.755
25	13.485	13.904	14.624	17.547	18.006	18.937	21.576	22.145	23.133	25.658	26.177	27.292	29.657	30.245	31.412
30	13.377	13.766	14.479	17.458	17.902	18.653	21.507	21.972	22.871	25.502	26.007	26.931	29.563	30.100	31.168
40	13.269	13.620	14.257	17.321	17.699	18.419	21.305	21.729	22.515	25.334	25.758	26.561	29.323	29.811	30.696
50	13.203	13.529	14.155	17.199	17.534	18.167	21.193	21.583	22.284	25.212	25.641	26.503	29.156	29.568	30.452
70	13.068	13.347	13.893	17.082	17.411	18.020	21.058	21.392	22.033	25.047	25.411	26.104	28.985	29.398	30.127
100	12.989	13.251	13.744	16.981	17.276	17.793	20.939	21.267	21.884	24.903	25.226	25.848	28.798	29.140	29.777

Notes: See Tables 1 and 2 for an explanation of the various features of this table.

Table 5: Size and size-adjusted power in Model 1 with no deterministic components.

			This study								Westerlund (2007)			
AR case	N	T	$\tau_{\hat{\alpha}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\overline{\tau}^*_{\hat{\alpha}_1}$	$\overline{\tau}^*_{\tilde{\alpha}_1}$	$\overline{w}^*_{\hat{\delta}_1}$	$\overline{w}^*_{\tilde{\delta}_1}$	G_τ	G_ρ	P_τ	P_ρ
							Size							
1	10	100	5.7	6.1	7.5	7.6	5.4	7.1	12.0	15.5	9.5	5.8	11.8	17.0
	10	200	5.3	5.5	6.1	6.3	5.5	6.1	8.0	11.2	7.9	6.1	11.1	16.1
	20	100	5.6	5.8	7.6	7.6	4.4	5.8	14.3	19.5	7.6	4.4	8.0	11.1
	20	200	5.4	5.6	6.2	6.4	4.9	7.1	9.5	14.9	5.8	4.0	8.1	11.2
2	10	100	6.1	6.1	7.9	8.4	6.7	6.6	13.3	18.1	2.1	0.9	2.9	3.4
	10	200	5.6	5.1	6.3	6.5	5.9	5.1	8.3	11.7	2.1	1.2	2.5	3.0
	20	100	6.1	6.0	8.0	8.4	5.7	6.3	16.9	22.8	2.2	0.9	2.6	3.7
	20	200	5.6	5.3	6.4	6.7	5.5	5.2	10.5	16.2	1.4	0.7	2.7	3.1
3	10	100	5.1	5.5	23.3	18.6	0.7	1.9	68.1	52.1	4.1	1.7	2.3	1.9
	10	200	5.0	4.8	12.1	9.3	2.7	2.7	28.0	19.2	5.2	2.0	7.3	7.9
	20	100	5.1	5.5	23.7	18.3	0.3	0.6	86.4	69.5	4.0	0.7	1.8	1.6
	20	200	5.0	4.8	12.0	9.1	1.5	2.0	37.7	23.3	4.2	0.6	6.2	5.2
4	10	100	7.0	6.4	8.7	12.3	8.4	6.3	16.0	31.1	9.8	6.6	12.0	17.5
	10	200	5.9	5.3	6.5	10.0	6.7	5.3	9.5	22.6	8.1	6.4	11.0	16.3
	20	100	6.8	6.0	8.9	12.8	7.9	5.5	22.1	38.9	7.9	4.4	8.1	11.5
	20	200	5.9	5.5	6.7	10.6	6.7	5.0	12.6	30.9	6.0	4.2	7.6	11.0

Continued overleaf

Table 5: Continued.

AR case	N	T	This study								Westerlund (2007)			
			$\tau_{\hat{\alpha}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\overline{\tau}^*_{\hat{\alpha}_1}$	$\overline{\tau}^*_{\tilde{\alpha}_1}$	$\overline{w}^*_{\hat{\delta}_1}$	$\overline{w}^*_{\tilde{\delta}_1}$	G_τ	G_ρ	P_τ	P_ρ
			Size-adjusted power											
1	10	100	16.4	13.0	28.4	13.1	77.6	60.8	85.9	39.0	28.7	20.3	34.3	45.9
	10	200	49.7	35.6	58.2	30.3	100.0	99.5	99.9	88.4	47.7	45.7	50.9	62.3
	20	100	16.8	13.6	28.4	13.8	95.6	81.8	97.1	50.8	54.5	41.6	55.2	62.8
	20	200	49.9	36.0	58.5	31.2	100.0	100.0	100.0	97.1	76.3	74.1	69.0	74.1
2	10	100	15.8	13.1	28.0	12.4	74.7	58.0	85.3	37.5	4.1	2.5	5.7	7.2
	10	200	48.8	33.9	57.7	29.5	100.0	98.2	99.9	86.9	6.6	5.4	7.3	11.1
	20	100	16.0	13.7	27.9	13.2	94.2	76.6	96.4	48.9	20.0	12.7	21.9	28.6
	20	200	49.2	34.5	58.1	30.3	100.0	99.8	100.0	96.9	34.1	32.2	33.1	41.5
3	10	100	12.0	9.9	6.8	3.5	56.2	44.8	14.7	4.8	25.9	13.1	11.6	14.2
	10	200	30.2	23.9	16.4	10.1	98.7	96.1	74.9	40.1	70.9	58.3	57.7	64.2
	20	100	12.5	10.0	6.6	3.6	79.8	68.9	22.0	6.2	67.5	36.9	19.5	19.7
	20	200	30.9	24.1	17.1	11.0	100.0	99.8	94.9	60.6	96.8	91.4	87.4	88.5
4	10	100	25.8	15.5	56.8	22.4	90.7	58.8	99.4	51.2	18.9	13.9	22.6	31.7
	10	200	71.1	46.8	82.6	43.2	100.0	99.0	100.0	91.8	29.2	28.5	32.2	42.2
	20	100	26.6	16.4	56.5	22.6	98.9	76.4	100.0	57.7	32.0	22.7	32.7	38.9
	20	200	71.3	47.6	82.6	43.4	100.0	99.9	100.0	98.1	47.8	47.3	43.0	47.6

Notes: AR case 1 refers to the setup with no serial correlation, while cases 2 to 4 refer to the setup with serial correlation in the errors driving $\Delta Y_{i,t}$, $\Delta X_{i,t}$ and ΔF_t, respectively. In all cases, the serial correlation is of the first-order AR type with a common AR parameter of magnitude 0.5. The error correction parameter α_1 is zero under the null and -0.05 under the alternative.

Table 6: Size and size-adjusted power in Model 2 with an unrestricted constant.

			This study								Westerlund (2007)			
AR case	N	T	$\tau_{\hat{\alpha}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\overline{\tau}^*_{\hat{\alpha}_1}$	$\overline{\tau}^*_{\tilde{\alpha}_1}$	$\overline{w}^*_{\hat{\delta}_1}$	$\overline{w}^*_{\tilde{\delta}_1}$	G_τ	G_ρ	P_τ	P_ρ
							Size							
1	10	100	6.2	6.3	8.1	8.2	6.1	7.4	13.4	16.0	12.5	6.4	13.2	14.9
	10	200	5.7	5.7	6.5	6.4	6.0	6.9	8.6	11.7	9.1	6.5	11.8	14.2
	20	100	6.0	6.2	8.1	8.0	5.6	7.3	15.9	20.5	12.2	4.9	10.5	10.0
	20	200	5.5	5.7	6.5	6.6	5.9	7.8	9.8	15.0	9.0	5.1	10.1	10.4
2	10	100	6.9	6.3	8.6	9.1	7.9	7.9	15.5	20.3	4.4	1.0	4.0	1.2
	10	200	6.0	5.3	6.8	7.0	6.9	5.7	9.4	13.1	2.6	0.8	2.9	1.1
	20	100	6.7	6.3	8.7	8.9	7.9	7.7	19.7	26.1	4.8	1.1	3.9	2.4
	20	200	5.9	5.2	6.8	7.0	7.1	6.1	11.3	18.0	2.5	0.5	3.0	1.3
3	10	100	4.7	5.3	19.3	16.3	0.6	1.1	50.1	41.2	6.6	1.3	3.3	1.1
	10	200	5.1	4.8	11.1	8.7	2.5	2.5	22.1	15.2	7.0	1.9	7.9	4.5
	20	100	4.7	5.1	19.6	16.1	0.1	0.4	68.9	52.7	7.3	0.6	2.5	0.8
	20	200	5.0	4.6	10.8	8.4	1.5	1.8	27.3	18.4	5.9	0.5	7.1	2.0
4	10	100	7.8	6.5	9.7	11.2	10.6	6.5	20.2	25.8	12.7	6.6	14.1	15.4
	10	200	6.5	5.4	7.3	8.8	7.8	5.5	11.3	19.2	9.2	6.7	11.7	14.0
	20	100	7.7	6.4	9.8	11.2	12.2	6.6	26.2	32.0	12.9	5.3	11.4	10.9
	20	200	6.3	5.4	7.2	9.2	8.7	5.5	13.8	24.6	8.4	4.8	9.2	9.4

Continued overleaf

Table 6: Continued.

AR case	N	T	$\tau_{\hat{\alpha}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\overline{\tau}^*_{\hat{\alpha}_1}$	$\overline{\tau}^*_{\tilde{\alpha}_1}$	$\overline{w}^*_{\hat{\delta}_1}$	$\overline{w}^*_{\tilde{\delta}_1}$	G_τ	G_ρ	P_τ	P_ρ
			This study								Westerlund (2007)			
			Size-adjusted power											
1	10	200	9.1	8.7	17.6	10.1	31.7	26.7	60.5	25.1	29.9	21.3	34.2	43.5
	10	200	32.2	23.0	45.7	24.1	98.5	88.7	99.5	75.7	48.4	51.1	52.9	67.0
	20	100	9.4	8.6	18.0	10.3	49.5	39.4	82.2	33.8	55.2	45.2	56.4	65.5
	20	200	32.7	23.3	46.7	24.5	100.0	98.3	100.0	90.5	79.8	82.8	76.3	83.2
2	10	100	8.7	9.0	17.1	9.4	29.4	25.1	59.4	23.2	6.5	2.3	6.5	3.5
	10	200	31.5	22.4	45.6	22.9	98.4	85.7	99.6	73.9	7.4	4.8	7.3	6.9
	20	100	9.0	9.0	17.4	9.9	45.9	36.3	80.9	31.0	22.8	14.0	22.4	25.2
	20	200	32.0	23.1	46.3	23.4	99.9	96.2	100.0	88.4	34.8	33.7	33.7	41.9
3	10	100	8.4	7.3	7.4	5.0	25.3	22.0	15.6	8.5	21.5	11.0	7.5	6.7
	10	200	19.8	15.6	15.2	9.8	84.8	75.1	68.0	34.9	60.5	53.6	50.5	55.2
	20	100	8.8	7.5	7.2	5.1	41.6	34.0	23.5	11.1	54.1	28.1	17.4	17.5
	20	200	20.2	16.4	15.7	10.3	97.7	93.3	90.1	54.5	95.6	90.8	84.9	85.2
4	10	100	12.4	8.7	40.5	18.3	41.3	23.6	94.6	41.9	23.4	16.0	25.4	30.9
	10	200	53.6	31.9	74.2	40.2	99.8	89.4	100.0	88.9	32.6	32.6	35.3	44.4
	20	100	13.1	8.8	40.7	18.9	61.7	31.5	99.3	49.5	38.0	26.4	37.4	41.5
	20	200	54.3	32.7	74.6	40.1	100.0	97.1	100.0	96.1	52.6	53.9	48.3	52.2

Notes: See Table 5 for an explanation of the various features of this table.

Table 7: Size and size-adjusted power in Model 3 with unrestricted constant and trend terms.

			This study								Westerlund (2007)			
AR case	N	T	$\tau_{\hat{\alpha}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\overline{\tau}^*_{\hat{\alpha}_1}$	$\overline{\tau}^*_{\tilde{\alpha}_1}$	$\overline{w}^*_{\hat{\delta}_1}$	$\overline{w}^*_{\tilde{\delta}_1}$	G_τ	G_ρ	P_τ	P_ρ
							Size							
1	10	100	6.6	6.8	8.4	8.4	6.3	8.2	14.4	17.8	16.6	5.4	14.9	11.2
	10	200	5.9	5.9	6.6	6.5	6.1	7.4	9.2	12.6	9.8	6.0	11.6	10.9
	20	100	6.4	6.6	8.7	8.5	5.8	8.8	19.0	23.7	15.0	3.3	11.1	6.6
	20	200	5.8	6.0	6.8	6.9	6.2	9.2	11.7	18.4	8.7	3.6	8.3	6.5
2	10	100	7.3	6.7	9.1	9.7	8.5	7.5	17.2	22.7	3.8	0.7	1.7	0.4
	10	200	6.3	5.4	6.9	7.2	7.2	5.3	10.5	15.0	1.7	0.2	0.9	0.2
	20	100	7.2	6.6	9.4	9.7	9.0	8.8	24.2	31.6	4.2	0.9	2.1	0.6
	20	200	6.3	5.4	7.2	7.4	7.5	6.7	13.6	21.8	1.2	0.2	1.0	0.3
3	10	100	4.3	5.1	15.7	14.1	0.5	1.0	34.8	30.3	6.9	1.2	2.2	1.0
	10	200	5.1	4.7	9.7	7.8	1.8	2.1	16.6	12.8	5.7	1.6	5.7	2.7
	20	100	4.3	5.0	16.1	14.2	0.1	0.2	50.3	42.6	7.2	0.5	2.3	0.6
	20	200	5.1	4.5	10.0	8.1	1.4	1.6	22.6	17.5	4.8	0.2	5.1	0.6
4	10	100	8.4	6.9	10.4	10.2	12.3	6.9	23.7	22.4	17.0	6.2	15.9	12.0
	10	200	6.9	5.4	7.5	7.7	8.5	5.1	12.1	15.6	10.3	5.5	12.0	10.7
	20	100	8.4	6.7	10.7	10.3	14.1	7.8	35.2	28.3	16.8	4.1	11.6	6.9
	20	200	6.8	5.5	7.8	8.1	10.0	6.4	17.0	21.5	8.1	3.4	7.6	5.8

Continued overleaf

Table 7: Continued.

AR case	N	T	This study $\tau_{\hat{\alpha}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\overline{\tau}^*_{\hat{\alpha}_1}$	$\overline{\tau}^*_{\tilde{\alpha}_1}$	$\overline{w}^*_{\hat{\delta}_1}$	$\overline{w}^*_{\tilde{\delta}_1}$	Westerlund (2007) G_τ	G_ρ	P_τ	P_ρ
			Size-adjusted power											
1	10	200	6.2	6.4	11.7	8.0	10.5	12.3	37.1	16.3	26.3	11.5	26.2	22.4
	10	200	18.5	14.1	35.5	19.1	80.2	61.2	96.4	59.4	39.7	32.3	44.9	50.3
	20	100	6.4	6.6	11.9	8.2	14.9	15.6	53.5	19.3	40.7	15.9	35.7	28.3
	20	200	18.8	14.5	36.0	19.2	96.9	81.1	99.8	73.4	69.1	61.1	69.5	72.0
2	10	100	6.1	6.6	11.5	7.4	9.6	13.9	36.0	14.3	4.6	1.3	2.4	0.8
	10	200	17.7	14.6	35.0	17.5	78.7	58.9	96.3	55.0	4.4	2.0	2.9	1.6
	20	100	6.2	6.9	11.6	7.7	13.0	16.2	52.3	16.3	13.9	3.8	8.9	4.9
	20	200	18.2	15.0	35.8	18.0	96.5	75.2	99.8	70.2	19.8	12.4	18.6	17.5
3	10	100	6.5	6.1	7.3	5.9	13.0	10.1	14.3	9.1	13.1	3.5	3.7	2.3
	10	200	12.9	11.1	13.9	9.3	54.3	45.4	55.7	29.3	39.3	20.8	36.8	31.5
	20	100	6.7	6.2	7.3	5.8	17.3	14.0	20.4	11.1	28.5	3.9	11.1	4.9
	20	200	13.1	11.5	14.1	9.4	76.5	65.8	78.6	42.2	79.9	43.0	77.8	67.3
4	10	100	5.7	5.4	27.4	14.4	6.8	7.2	79.3	31.2	23.4	9.8	21.8	18.0
	10	200	33.2	18.4	65.6	36.7	95.3	60.5	100.0	82.4	27.6	20.5	30.2	32.8
	20	100	5.9	5.5	27.6	14.9	8.1	7.0	92.1	34.8	29.8	10.9	25.1	19.0
	20	200	33.8	18.6	65.6	36.0	99.8	75.2	100.0	91.9	42.3	33.1	40.0	39.5

Notes: See Table 5 for an explanation of the various features of this table.

Table 8: Cointegration test results for the Fisher effect.

	Model 2				Model 3			
Country	$\tau_{\hat{\alpha}^{\star}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}^{\star}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\tau_{\hat{\alpha}^{\star}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}^{\star}_{1i}}$	$w_{\tilde{\delta}_{1i}}$
Australia	−3.43*	−4.22**	5.19	13.95	−2.01	−2.82	17.77*	18.34*
Belgium	−1.84	−3.36	6.40	12.98	−2.82	−3.69	10.80	15.63
Canada	−2.33	−4.51***	9.16	22.38***	−3.26	−4.10*	15.89	24.82***
Switzerland	−0.70	−2.18	3.74	5.72	−1.09	−1.85	8.51	5.50
Germany	−2.93	−2.27	6.25	7.69	−3.30	−2.70	8.61	7.92
Denmark	−4.90***	−5.55***	28.43***	28.43***	−5.06***	−4.69**	32.46***	32.46***
Spain	−3.72*	−4.57***	14.57	22.46***	−4.48**	−4.62**	21.16**	23.09**
Finland	−2.74	−3.43***	9.97	13.27	−2.89	−3.32	10.83	13.04
France	−1.82	−2.54	7.49	7.70	−2.12	−3.18	11.01	12.95
United Kingdom	−4.39***	−4.57***	24.07***	25.54***	−4.41**	−4.36**	23.28**	24.89**
Ireland	−6.40***	−4.46***	28.43***	28.43***	−6.60***	−4.45**	32.46***	32.46***
Italy	−3.14	−4.16**	18.84**	25.10***	−3.72	−3.82*	22.96**	24.38**
Japan	−2.83	−5.60***	24.39***	28.43***	−3.30	−5.76***	22.40**	32.46***
Luxembourg	−2.83	−4.18**	11.46	23.01***	−3.84*	−4.37*	18.35*	23.89**
Netherlands	−0.31	−2.39	4.57	10.96	−0.97	−2.89	6.19	13.17
Norway	−2.70	−3.00	8.83	11.71	−4.15*	−3.93*	8.67	11.23
New Zealand	−6.40***	−6.40***	28.43***	28.43***	−6.14***	−6.19***	32.46***	32.46***
Portugal	−3.27	−3.47*	28.43***	26.94***	−3.46	−3.20	32.18***	25.26***
Sweden	−2.49	−4.44***	13.77	22.08***	−3.82*	−4.49*	23.40**	22.28**
United States	−1.46	−1.72	2.99	5.19	−6.06***	−6.32***	32.46***	32.46***
All 20[a]	−3.03***	−3.85***	14.27***	18.52***	−3.68***	−4.04***	19.59***	21.43***
	G_τ	G_ρ	P_τ	P_ρ	G_τ	G_ρ	P_τ	P_ρ
All 20[b]	−8.74***	−9.30***	−7.43***	−10.01***	−8.72***	−6.64***	−7.00***	−6.12***

Notes: The superscripts (***), (**) and (*) denote significance at the 1%, 5% and 10% levels, respectively. The superscript ($\star$) in $\tau_{\hat{\alpha}^{\star}_{1i}}$ and $w_{\hat{\delta}^{\star}_{1i}}$ indicates that these tests have been computed using principal component estimates of the factors instead of cross-sectional averages. See Table 1 for an explanation of the two models.

[a]The reported values are for the average of the truncated test statistics.

[b]The reported values are obtained by applying the Westerlund (2007) tests to the defactored data.

Table 9: Cointegration test results for the monetary exchange rate model.

	Model 2						Model 3					
Country	$\tau_{\hat{\alpha}^{\star}_{1i}}$	$\tau_{\hat{\alpha}^{\star\star}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}^{\star}_{1i}}$	$w_{\hat{\delta}^{\star\star}_{1i}}$	$w_{\tilde{\delta}_{1i}}$	$\tau_{\hat{\alpha}^{\star}_{1i}}$	$\tau_{\hat{\alpha}^{\star\star}_{1i}}$	$\tau_{\tilde{\alpha}_{1i}}$	$w_{\hat{\delta}^{\star}_{1i}}$	$w_{\hat{\delta}^{\star\star}_{1i}}$	$w_{\tilde{\delta}_{1i}}$
United Kingdom	−2.38	−3.19	−1.41	12.70	20.36*	12.16	−2.32	−3.60	−1.48	12.76	27.75**	13.71
Austria	−2.54	−2.62	−1.59	9.59	10.88	17.46	−1.62	−2.49	−2.46	6.68	10.28	16.22
Belgium	−4.62**	−2.60	−1.63	23.26**	10.10	15.94	−4.72**	−3.05	−1.53	27.91**	12.57	18.45
Denmark	−2.20	−2.66	−1.03	15.81	23.68**	16.04	−3.07	−2.18	−1.14	21.02	23.12*	16.48
France	−1.58	−2.87	−1.98	12.26	14.23	6.55	−2.52	−2.36	−3.23	14.14	14.12	13.25
Germany	−2.11	−3.10	−2.23	9.27	12.63	23.19**	−3.84	−3.18	−2.58	19.92	12.63	22.46*
Netherland	−2.66	−2.53	−1.04	10.15	9.19	15.33	−2.54	−2.58	−2.38	8.88	11.11	19.04
Canada	−2.46	−1.16	−2.94	9.30	11.56	17.31	−2.56	−1.38	−2.89	9.87	9.47	17.57
Japan	−4.06**	−2.38	−0.67	22.17*	17.44	9.36	−2.68	−2.33	−1.23	22.24	16.47	10.07
Finland	−1.86	−2.29	−2.56	19.95*	7.40	16.49	−2.67	−2.83	−2.40	24.41*	11.77	17.07
Greece	−2.89	−3.53	−2.55	14.67	16.69	14.40	−2.99	−3.51	−3.60	15.14	17.83	21.51
Spain	−4.46**	−3.71	−5.84***	24.88**	31.11***	36.09***	−4.49**	−3.76	−5.80***	25.78**	31.28***	39.69***
Australia	−3.27	−2.90	−3.40	15.87	12.07	20.78*	−3.44	−2.91	−2.82	15.63	11.28	19.11
Italy	−1.99	−2.19	−4.54**	13.33	17.32	23.24**	−3.54	−2.00	−4.62**	23.12*	15.73	23.54*
Switzerland	−2.22	−3.55	−3.05	10.26	15.52	36.09***	−2.07	−3.42	−2.77	5.86	12.89	32.35***
Korea	−1.07	−3.79	−3.37	30.07***	36.09***	35.01***	−1.82	−4.11	−3.47	28.10**	38.30***	32.46***
Norway	−3.20	−2.62	−4.48**	18.79	10.48	29.89***	−2.96	−2.60	−4.39*	18.53	10.33	30.37**
Sweden	−1.68	−2.26	−4.17**	6.91	15.54	23.20**	−2.74	−2.33	−4.15	11.36	15.51	22.95*
European community[a]	−2.73*	−2.93**	−2.50	16.38***	17.26***	15.90***	−3.17**	−2.91	−3.00	19.29***	18.46***	18.61***
European monetary system[b]	−1.59	−2.70*	−2.42	13.84**	14.53**	13.92**	−2.67	−2.56	−3.04	18.82***	14.88	15.21*
Group of 6[c]	−2.42	−2.48	−1.98	14.31**	15.59***	12.28	−2.56	−2.48	−2.45	16.32**	16.03**	14.78
Group of 10[d]	−2.48	−2.58	−2.48	14.10**	14.39**	15.18***	−2.71	−2.62	−3.19**	14.90	14.82	19.86***
All 18	−2.63*	−2.78**	−2.69**	15.51***	16.24***	20.48***	−2.92	−2.81	−2.94	17.30***	16.80***	21.46***

Notes: The individual tests are based on the full panel averages. The superscript (*) in $\tau_{\hat{\alpha}^{\star}_{1i}}$ and $w_{\hat{\delta}^{\star}_{1i}}$ indicates that these tests have been computed using the observed factors, while (**) indicates that the tests are based on applying the principal components method to a pre-specified cointegrating relationship. See Table 8 for an explanation of the remaining features of the table.

[a] Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Spain and United Kingdom.

[b] Belgium, Denmark, France, Germany, Italy, Netherlands, Spain.

[c] Canada, France, Germany, Italy, Japan and United Kingdom.

[d] Belgium, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland and United Kingdom.

Table 10: Descriptive statistics for the common factors in the monetary exchange rate model.

	Principal components			Observed	
Value	Factor 1	Factor 2	Factor 3	m_t^*	y_t^*
AR	0.97	0.98	0.97	1.00	1.00
SE	0.03	0.05	0.02	0.00	0.01
ADF	−1.09	−0.34	−1.79	-3.06^{**}	−0.42

Notes: AR refers to the estimated first order AR coefficient, SE refers to its standard error and ADF refers to the augmented Dickey and Fuller (1979) test. The autoregressions are fitted with an intercept and the lag orders are determined using the Schwarz Bayesian criterion. See Table 8 for an explanation of the remaining features of the table.